

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 27, 2002, 13:45:31 ; Search time 39 Seconds
(without alignments)
1701.508 Million cell updates/sec

Title: US-09-926-084-7
Perfect score: 2601
Sequence: 1 MDLSLFVSVAVLVGSSSHV.....KCGEYFFLSALTATIAA 498

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: A_Geneseq.101002.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2581	99.2	498	21	AA823718
2	185	7.1	421	21	AAV75546
3	177	6.8	421	21	AAV75545
4	174	6.7	421	21	AAV75544
5	137.5	5.3	439	22	AAU36119
6	130.5	5.0	423	22	AAU45528
7	121.5	4.7	1093	23	ABBS4302
8	121	4.7	508	22	AAU44925
9	118	4.5	2383	21	AA815945
10	114.5	4.4	567	22	ABG28940

11	112.5	4.3	1837	21	AA811726
12	112.5	4.3	1837	23	ABJ04044
13	112	4.3	197	22	ABG28941
14	111.5	4.3	523	22	AA895114
15	111.5	4.3	1194	22	ABG23389
16	110	4.2	21	21	AA823715
17	110	4.2	1909	22	ABG19127
18	109.5	4.2	421	23	ABG48147
19	109.5	4.2	4924	22	AA870968
20	109.5	4.2	4928	20	AAJ39300
21	109	4.2	1600	22	AA844976
22	107	4.1	21	21	AA823714
23	107	4.1	940	22	ABBA7334
24	106.5	4.1	615	22	AA892313
25	105.5	4.1	729	22	ABG04814
26	105	4.0	753	22	AAU34588
27	105	4.0	2469	22	ABG10456
28	105	4.0	3386	22	ABG29101
29	104.5	4.0	1684	12	AA814948
30	104	4.0	603	14	AA844004
31	102.5	3.9	902	21	AA812995
32	102.5	3.9	1003	8	AA870861
33	102.5	3.9	1712	22	ABBA4566
34	102	3.9	513	9	AA81501
35	102	3.9	2089	17	AAW08333
36	101.5	3.9	1003	7	AA860420
37	101	3.9	398	22	AAU33973
38	101	3.8	409	22	AAU36825
39	100	3.8	664	22	ABG04472
40	100	3.8	664	22	ABG05550
41	100	3.8	708	21	AAV75566
42	100	3.8	905	22	ABG05208
43	100	3.8	1022	9	AA81854
44	99	3.8	1389	21	AAV74854
45	98.5	3.8	712	19	AAW53110

ALIGNMENTS

RESULT 1	AA823718	standard; Protein; 498 AA.
XX	AA823718	
AC	AA823718;	
XX	09-JAN-2001	(first entry)
XX	Geotrichum candidum Dec 1 (FERM BP-7033)	DYP protein SEQ ID NO:7.
DE	Geotrichum candidum Dec 1 (FERM BP-7033); DYP; decolouring activity;	
XX	Geotrichum candidum Dec 1 (FERM BP-7033); DYP; decolouring activity;	
KW	dye; peroxidase; dye-decomposition; waste; textile.	
OS	Galactomyces geotrichum.	
XX	WO2000050582-A1.	
PN	31-AUG-2000.	
XX	25-FEB-2000; 2000WO-JP01093.	
XX	26-FEB-1999; 99JP-0050562.	
XX	(MEIU) MEIU SEIKA KAISHA LTD.	
PA	Syoda M, Sugano Y, Kubota H;	
PI	WPI; 2000-549408/50.	
XX	N-PSDB; AAA92280.	
DR	Geotrichum candidum-originated peroxidase with high dye-decomposing	
XX	activity on wide range of dyes, useful for waste treatment in e.g.	
PT	dyeing and textile industries, without secondary contamination or	
PT		

PT consuming much energy -

XX -Claim 2; Page 36-39; 43pp; Japanese.

XX

CC The present sequence represents a peroxidase designated Dyp originating from Geotrichum candidum Dec 1 (FERM BP-7033), also known as Galactomycetes geotrichum. The peroxidase has dye-decomposing and decolouring activity, a molecular weight of 60 kDa by SDS-PAGE (sodium dodecyl sulphate), or 55 kDa (by gel filtration), and an isoelectric point of 3.8. The peroxidase is useful for waste treatment in the dyeing industry and textile industry. The enzyme has a high dye-decomposing activity on a wide range of dyes, but without secondary contamination or consuming much energy to damage to environment or give greenhouse gas.

XX

SQ Sequence 498 AA;

Query Match 99.2%; Score 2581; DB 21; Length 498;

Best Local Similarity 99.4%; Pred. No. 1.1e-223;

Matches 495; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MDLSLFVSVAVLVGSSSHVNAKLGARQRTTPLLTFNFGQAPLPLTQHTTESGANDT 60

DB 1 MRLSLFVSVAVLVGSSSHVNAKLGARQRTTPLLTFNFGQAPLPLTQHTTESGANDT 60

QY 61 ILPLNNTGGDILVGMKKOKERFVFFQNDATSKTALKTVVQRIITSAAILISDPSOQPL 120

DB 61 ILPLNNTGGDILVGMKKOKERFVFFQNDATSKTALKTVVQRIITSAAILISDPSOQPL 120

QY 121 AFVNLGFSNTGLQALGITDLDGDAQFPDGFADAAANLGGDLSQWVAPFTGTHGVFLIG 180

DB 121 AFVNLGFSNTGLQALGITDLDGDAQFPDGFADAAANLGGDLSQWVAPFTGTHGVFLIG 180

QY 181 SQDDFLDQTDIDISSTFGSSITQVQALSGSARPDQAGHEHFGFLDGIQSPSVTGWETT 240

DB 181 SQDDFLDQTDIDISSTFGSSITQVQALSGSARPDQAGHEHFGFLDGIQSPSVTGWETT 240

QY 241 VFGQAVVPGIILTGDDGTGTRPSWALDGSFMAFRHFOOKVPEFNAYLANAIPANSA 300

DB 241 VFGQAVVPGIILTGDDGTGTRPSWALDGSFMAFRHFOOKVPEFNAYLANAIPANSA 300

QY 301 GNLTQGEAEFLGARMGRKWSGAPIDLAPTADDPALGADPQRNNNFYSDTLTDETRCP 360

DB 301 GNLTQGEAEFLGARMGRKWSGAPIDLAPTADDPALGADPQRNNNFYSDTLTDETRCP 360

QY 361 FGARVKTNPQDLGGPVDFTFHAMRSSIPYGPETSDAELASGVTAQDRGLLFVEYQSIIG 420

DB 361 FGARVKTNPQDLGGPVDFTFHAMRSSIPYGPETSDAELASGVTAQDRGLLFVEYQSIIG 420

QY 421 NGRFQOINWANNANFPFSKPIITPGIEPIIGQTPRTVGGDPLNQNETVPLFVLPKG 480

DB 421 NGRFQOINWANNANFPFSKPIITPGIEPIIGQTPRTVGGDPLNQNETVPLFVLPKG 480

QY 481 GEYFFLPSISALTATIAA 498

DB 481 GEYFFLPSISALTATIAA 498

RESULT 2

AAV75546

ID AAV75546 standard; Protein; 421 AA.

XX

AC AAV75546;

XX

DT 21-MAR-2000 (first entry)

XX

DE Neisseria meningitidis ORF 748 protein sequence SEQ ID NO:2566.

XX

KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine; antigenic; diagnosis; immunogenic; infection; meningitis; septicemia; antibacterial; gene therapy.

XX

OS Neisseria meningitidis.

XX

PN WO9957280-A2.

XX

PD 11-NOV-1999.

XX

PF 30-APR-1999; 99WO-US09346.

XX

PR 01-MAY-1998; 98US-0083758.

PR 31-JUL-1998; 98US-0094869.

PR 02-SEP-1998; 98US-0098994.

PR 02-SEP-1998; 98US-0099062.

PR 09-OCT-1998; 98US-0103749.

PR 09-OCT-1998; 98US-0103794.

PR 09-OCT-1998; 98US-0103796.

PR 25-FEB-1999; 99US-0121528.

XX

PA (CHIR) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

XX

PI Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;

PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;

PI Tettelin H, Venter JC;

XX

DR WPI; 2000-062150/05.

DR N-PSDB; AA254308.

XX

DR Novel Neisserial polypeptides predicted to be useful antigens for vaccines and diagnostics

PT

PT

XX

PS Claim 2; Page 1219; 1453pp; English.

XX

XX AA253015 to AA254536, AA254577 to AA254615, and AA274253 to AA275941 represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent PCR primers used in the exemplification of the present invention. The polypeptides, the polynucleotides, antibodies and compositions of the invention can be used as vaccines, as diagnostic reagents, and as immunogenic compositions. The polypeptides can be used in the manufacture of medicaments for treating or preventing infection due to Neisserial bacteria (e.g. meningitis and septicemia), to detect the presence of Neisseria bacteria, or to raise antibodies. They may also be used to screen for agonists or antagonists, which may themselves have use as antibacterial agents. The polynucleotides of the invention may also be used in gene therapy protocols.

XX

SQ Sequence 421 AA;

Query Match 7.1%; Score 185; DB 21; Length 421;

Best Local Similarity 23.2%; Pred. No. 7e-08;

Matches 88; Conservative 41; Mismatches 125; Indels 126; Gaps 17;

QY 135 LGITDLDGDAQFPDQFA-----DAAN-----LGDDLQWVAPFTGTTIH 174

DB 130 LTVTVGVGSSSLF-DGRFGLKDKKPIHLQEMRDFSNBKLQKSCDGLSLQICAPTETCQ 188

QY 175 GVFLIGSDQDDFLDQFTDDISSTFGSSITQ--VQALSGSARPDQAGHEHFGFLDGIQSP 232

DB 189 AA-----LRDIKHTVQAVIRWSIDGWQPKSEPGAMARNLLGFRDGTGNP 235

QY 233 SVTGMETTVFPGQAVVPPGIIILTGDRDGTGTRPSWALDGSFMAFRHFOOKVPEFNAYTLA 292

DB 236 KVSDDPKTA-----DEVLTGVAANSLSDEPAWAKNGSYQAVLRIRHPVEFWDTPL- 285

QY 293 NAIPANSAGNLTOQGEAEFLGARMFGWKSGAPIDLAPTADDPALGADPQRNNNFYSDT 352

DB 286 -----QEQTDFIGRRKY----SGAPMDGKKEADQDFAKDPEGNTT----- 322

QY 353 LTDETRCFCAHVKTNPQDLGCP--VDTFHAMRSSIPYGPETSDAELASGVTAQDRGL 410

DB 323 -----PKDSHIRLANPRD-----PEFLKKHRLFRAYSF----SRGLASSG--QLDVLG 365

QY 411 LFVEYQSIIGNGRFQOINWANNANFPFSKPIITPGIEPIIGQTPRTVGGDPLNQNETF 470

DB 366 VFVCYQANLADGFIFVQ-NLLNG-----EPLLE----- 392

QY 471 TVPLFVLPKGEYFF-LPSI 489
 Db 393 ----YISPFGGYFFVLPGV 408

RESULT 3

AAV75545
 ID AAV75545 standard; Protein; 421 AA.

AC AAV75545;

DT 21-MAR-2000 (first entry)

DE Neisseria meningitidis ORF 748 protein sequence SEQ ID NO:2564.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 XX antibacterial; gene therapy.

XX Neisseria meningitidis.

PN WO957280-A2.

XX 11-NOV-1999.

PF 30-APR-1999; 99WO-US09346.

PR 01-MAY-1998; 98US-0083758.

PR 31-JUL-1998; 98US-0094869.

PR 02-SEP-1998; 98US-0098994.

PR 02-SEP-1998; 98US-009062.

PR 09-OCT-1998; 98US-0103749.

PR 09-OCT-1998; 98US-0103794.

PR 09-OCT-1998; 98US-0103796.

PR 25-FEB-1999; 99US-0121528.

DR N-PSDB; AAZ54307.

PI Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;

XX Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -

PS Claim 2; Page 1218; 1453pp; English.

XX AA253015 to AA254536, AA254577 to AA254615, and AAV74253 to AAV75941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.

XX Sequence 421 AA;

Query March 6.8%; Score 177; DB 21; Length 421;

Best Local Similarity 23.4%; Pred. No. 3.7e-07;
 Matches 89; Conservative 41; Mismatches 124; Indels 126; Gaps 18;

QY 135 LGITDLDGDAQFPDQFA-----DAAN-----LGDLDLQWVAPFTGTIY 174

Db 130 LTVTVGSSSLF-DGRGLDKKXPIHLQEMRDFSNDKLQKWCDDGLSLQICAFPTPTCQ 188
 QY 175 GVFLIGSDDDFLDQFPTDISSTFGSSITO--VQALSGSARPPDQACHFGLDGSOP 232
 Db 189 AA-----LNDIKHTVQTVAVIRWSIDGWQPKSEPCGMARNLGFRDGNP 235
 QY 233 SVTGMETTVPPGAVVBPPIILTGRDGTGRPSWALDGSFMAFRHPOKVPENAYTLA 292
 Db 236 KVSDEPKTA-----DEVLMTCVAAANSIDPEPMANSGVQAVRLIRHFEVFDRTPL- 285
 QY 293 NATPANSAGNLTQOEGAEFGARMFGRKSGAPRIDLAFTADDPALGADPQARNNFYSDT 352
 Db 286 -----OEQTDIGRRKY-----SGAPWDGKKEADQDPFAPKDP----- 318
 QY 353 LTBETRCPPGAVHVKTPRODLGSP--VDTFHMRSSIPYGPERSDMLASGVTAODRGL 410
 Db 319 -GDIT--PKOSHRLANPRD---PFLKXRLFRRAVSY---SRGLASG--QLDVGL 365
 QY 411 LFEVYQSIIGNGFRRFOQINNANNANFPFSKPITPGIEPIIGQTPPTVGGIDPLNQNETF 470
 Db 366 VFVCYQANLADGFIFFVQ-NLLNG-----EPLEF----- 392
 QY 471 TVPLFVLPKGEYFF-LPSI 489
 Db 393 ----YISPFGGYFFVLPGV 408

RESULT 4

AAV75544
 ID AAV75544 standard; Protein; 421 AA.

XX AAV75544;

DT 21-MAR-2000 (first entry)

DE Neisseria gonorrhoeae ORF 748 protein sequence SEQ ID NO:2562.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KW antibacterial; gene therapy.

XX Neisseria gonorrhoeae.

PN WO957280-A2.

XX 11-NOV-1999.

PF 30-APR-1999; 99WO-US09346.

PR 01-MAY-1998; 98US-0083758.

PR 31-JUL-1998; 98US-0094869.

PR 02-SEP-1998; 98US-0098994.

PR 02-SEP-1998; 98US-009062.

PR 09-OCT-1998; 98US-0103749.

PR 09-OCT-1998; 98US-0103794.

PR 09-OCT-1998; 98US-0103796.

PR 25-FEB-1999; 99US-0121528.

PI Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;

XX Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -

PS Claim 2; Page 1217-1218; 1453pp; English.

```
XX AA253015 to AA254536, AA254577 to AA254615, and AA274253 to AA275941
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisseria meningitidis (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria meningitidis, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 421 AA;
Query Match 6.7%; Score 174; DB 21; Length 421;
Best Local Similarity 23.0%; Pred. No. 6.9e-07;
Matches 87; Conservative 39; Mismatches 130; Indels 122; Gaps 15;
QY 135 LGITDLDGDAQFPDGPFA-----DAAN-----LGDLSQWVAPFTGTH 174
Db 130 LTVTVGVGSSLF-DGRFLGDKKXTHLQEMRDFPNDKLOKSWCDGLSLQICAFPTETCQ 188
QY 175 GVFLIGSDQDDFLDQDFTDSSITFGSSITO--VQALSGSARPPDQAGHEFGFLDISQP 232
Db 189 -----TALRDIHKTAQTAIVIRMSIDGMQPKSEPCAMARNLLGFRDGTGNP 235
QY 233 SVTGMETVFPQAVVPPGIIITGRDGDGTGRPSWALDGSFMAFRHPQOKVPEFNAYTLA 292
Db 236 KVSDPKTA-----DEVLTGTVAAANSLDEPEWAKNGSYQAVLLIRRVFVWDRTP- 285
QY 293 NAIPANSAGNLTOQEGAEFLGARMFGKWSGAPIDLAPTADDPALGADPQORNFFDYSDT 352
Db 286 -----QEQTDFGRKKY----SGAPMDGKKKADQPDFAKDPE----- 318
QY 353 LDTETRCFPGAHVKNRNPQDLGGPVDTFHAMRSSIPYGPETSDAELASGVTAQDRGLLF 412
Db 319 -GDIT--PKDSHMLANPRDPEFLKKHCLFRRAYSYSRGPASSG-----QLDVGLVF 367
QY 413 VEQSIIGNGFRFQOINWANNANFPFSKPTPGIEPIIGOTTPTRTVGGDLPLAQNETFTV 472
Db 368 VCFQANLADGFIQVQ-NLLMG-----BPLEE----- 392
QY 473 PLEVIPKGEYFF-LPSI 489
Db 393 --YISPPGGGYFFVLPGV 408
RESULT 5
AAU36119
ID AAU36119 standard; Protein; 439 AA.
XX AAU36119;
XX
XX 14-FEB-2002 (first entry)
XX Klebsiella pneumoniae cellular proliferation protein #107.
XX
XX
XX Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
XX Klebsiella pneumoniae.
XX
XX WO200170955-A2.
XX
XX 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US09180.
XX
XX 21-MAR-2000; 2000US-191078P.
XX
XX 23-MAY-2000; 2000US-206848P.
```

```
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253825P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlseen KL, Zyskind JW, Wall D, Trawick JD, Carr GU;
XX Yamamoto RT, Xu HH;
XX WPI: 2001-611495/70.
XX DR N-PSDB; AAS53978.
XX
XX New polynucleotides for the identification and development of
XX PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 11712; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
XX genes, their use in the discovery of novel antibiotics, the essential
XX genes themselves and the encoded proteins. The prokaryotes used are
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX invention is also useful for the identification of potential new targets
XX for antibiotic development. The antisense nucleic acids can also be used
XX to identify proteins used in proliferation, to express these proteins,
XX and to obtain antibodies capable of binding to the expressed proteins.
XX The proteins can be used to screen compounds in rational drug discovery
XX programmes. The antisense nucleic acid sequence is also useful to screen
XX for homologous nucleic acids which are required for cell proliferation in
XX a wide variety of organisms. The present sequence represents an
XX essential prokaryotic cellular proliferation protein.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 439 AA;
Query Match 5.3%; Score 137.5; DB 22; Length 439;
Best Local Similarity 21.2%; Pred. No. 0.0014;
Matches 108; Conservative 47; Mismatches 160; Indels 195; Gaps 22;
QY 4 SLFWVSVAVLVGSSSHVNAKLGARQTRTPTLLTNPPGQAP-----LPTLQHTTSGA 57
Db 87 SMLVAFDVLAAADKADLE--RLFRLLTQRIAFLTQ--GGPAPDTPNPRLP- 138
QY 58 NDTILPLNNIOGDILVGMKKOKERFVFFQVNDATSEKTKYVVPORITSAAILISDPQ 117
Db 139 LGPWIAPDNLTITVSVGHSLFDERF-----GLADKAPKKL----- 173
QY 118 QPLAFVNLGFSNTGLQALGITDLDGDAQFQDGADAAALGDDLSQWVAPFTGTHGVF 177
Db 174 QPW-----TRFNDSLDAALCHGDLLLQICANTOOTVIAHR 210
QY 178 LTGSDQDDFL-----DQFTDDISSTFGSSITQVQALSGSARPPDQAGHEFGFLDISQ 231
Db 211 DVIKHTPOLLVSVRWKREGFISDAARSKKETPINLL-----GFKDGTAN 255
QY 232 PSVTGWETTVFPQAVVPPGIIITGRDGDGTGRPSWALDGSFMAFRHPQOKVPEFNAYTL 291
Db 256 PA-----SHDSALMDKVMVVTADQD---EPTTWGGSYQAARIIQPHV-EFMDRT- 301
QY 292 ANNAIPANSAGNLTOQEGAEFLGARMFGKWK--SGAPIDLAPTADDPALGADPQORNFFDYS 350
Db 302 -----PLKEQQ-----TIFGRDKHTGAPLGMKNEHDT- 332
QY 351 DTLTDETRCFFGAHVKNRNPQDLGGPVDTFHAMRSSIPYGPETSDA-----ELASGV 403
Db 333 KOPNGEV-IALDSHIRLANPRT-----PETOSSLMRRGYSYSLGV 372
```

QY 404 T---AQRGLFVEYQSIIGNGFRFOQIMWANNANFPSPKPTIGIEPIIGQTPRTVGG 460
DB 373 TNAGQLMGLLFVICYQHDLKGFVTVOKRLNGEALKEEVKPI----- 414
QY 461 LDPINQNETFTVPLFVLPKGEYFF-LPSI 489
DB 415 -----GGGYFPVLPGV 425

RESULT 6
AAU34528
ID AAU34528 standard; Protein; 423 AA.
XX
AC AAU34528;
XX
DT 14-FEB-2002 (first entry)
XX
DE E. coli cellular proliferation protein #109.
XX
KM Antisense; prokaryotic cellular proliferation protein;
KM antibiotic; antibacterial; drug design.
XX
OS Escherichia coli.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
XX
PR 23-MAY-2000; 2000US-206848P.
XX
PR 26-MAY-2000; 2000US-207272P.
XX
PR 23-OCT-2000; 2000US-242578P.
XX
PR 27-NOV-2000; 2000US-253625P.
XX
PR 22-DEC-2000; 2000US-257931P.
XX
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
XX
N-PSDB; AAS53387.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX
PS Example 3; Seq ID No 10121; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 423 AA;

Query Match 5.0%; Score 130.5; DB 22; Length 423;
Best Local Similarity 22.3%; Pred. No. 0.0058;
Matches 81; Conservative 29; Mismatches 119; Indels 135; Gaps 14;

QY 145 QFPDGGPADAANIGDLSQVAPFTGTTTIGVGLISDDDDPL-----DQFTDDISSTF 128
DB 162 RFPVDSIDAALCHGDVLLQICANTQDTVHALRDILIKRPDILSVRKRGFSIDHAARS 221
QY 199 GSSITQVALSGSARPPDQAGHHEFGFLDGISQPSVTG-----WETVFPQAVVP 250
DB 222 KGETPIINL-----GFKDGTANPDSQNDKLMQKVVVTA----- 256
QY 251 GIIVTGRDGDGTGRPSWALDSFMAFRHQKVPFBNVAYTLANAIPANSAGNLTOEGAE 310
DB 257 -----DQGEPAWTIGSGYQAVVLIQFRV-EFWDRT-----PLKEQD--- 291
QY 311 FLGARMGRMK-SGAPIDLAPTADDPALGADPQARNNFYDILTDETRCPFGAHVAKTN 369
DB 292 -----TIFGRDKQTGAPLGMQHEHDVPYASDPGK-----VIALDSHRLAN 334
QY 370 PRODLGGPVDTFPMARSSIFYPETSDAELASGVTAQ---DRGLLFVEYQSIIGNGFRFQ 426
DB 335 PRT---AESSSIMLRGTY-----SLGVTNSGQIDMKGLLFVICYQHDLKGFPLTV 382
QY 427 QIMWANNANFPSPKPTIGIEPIIGQTPRTVGGDPLANQNETFTVPLFVLPKGEYFF- 485
DB 383 QKRLNGEALKEEVKPI-----GGGYFFA 405
QY 486 LPSI 489
DB 406 LPGV 409

RESULT 7
ABB54302
ID ABB54302 standard; Protein; 1093 AA.
XX
AC ABB54302;
XX
DT 16-MAY-2002 (first entry)
XX
DE Lactococcus lactis protein ykbc.
XX
KM Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
XX
OS Lactococcus lactis IL1403.
XX
PN FR2807446-A1.
XX
PD 12-OCT-2001.
XX
PF 11-APR-2000; 2000FR-0004630.
XX
PR 11-APR-2000; 2000FR-0004630.
XX
PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
XX
PI Bolotine A, Sorokine A, Renault P, Ehrlich SD;
XX
DR WPI; 2002-043418/06.
XX
PT New nucleotide sequence useful in the identification or Lactococcus
PT lactis and related species -
XX
XX
PS Claim 6; SEQ ID No 1004; 2504pp; French.
XX
XX The present invention is related to a Lactococcus lactis nucleotide
CC sequence (ABB54302) and related proteins (ABB53300-ABB55621). The
CC nucleic acid sequence is useful in the detection and/or amplification of
CC nucleic acid sequence, particularly to identify Lactococcus lactis or
CC related species. The proteins of the invention are useful for the
CC biosynthesis or biodegradation of a composition of interest. The
CC invention helps research in lactic bacteria, particularly useful in the

AAB15945
 ID AAB15945 standard; Protein; 2383 AA.
 XX
 AC AAB15945;
 DT 05-OCT-2000 (first entry)
 XX
 DE E. coli proliferation associated protein sequence SEQ ID NO:302.
 XX
 KM Escherichia coli; E. coli; proliferation; inhibition; screening;
 XX antimicrobial; bacterial growth; antisense therapy; antibacterial.
 OS Escherichia coli.
 XX
 PN WO200044906-A2.
 XX
 PD 03-AUG-2000.
 XX
 PF 27-JAN-2000; 2000WO-US02200.
 XX
 PR 27-JAN-1999; 99US-0117405.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Zyskind J, Ohlsen KL, Trawick J, Forsyth RA, Froelich JM, Carr GJ,
 PI Yamamoto RT, Xu HH;
 XX WPI; 2000-514822/46.
 DR N-PSDB; AAB65950.
 XX
 PS Novel polynucleotides and polypeptides associated with microorganism
 PT proliferation, used to identify inhibitors of bacterial growth and
 PT proliferation, for use in antisense therapy -
 XX
 PS Claim 11; Page 224-229; 316pp; English.
 XX
 CC AAB65809 to AAB65889 and AAB66058 to AAB66138 represent nucleotide
 CC sequences derived from Escherichia coli which inhibit E. coli
 CC proliferation. AAB65890 to AAB66055 and AAB15886 to AAB16040 represent
 CC nucleotide and protein sequences associated with E. coli proliferation.
 CC AAB66056 and AAB66057 represent primers used for sequencing E. coli
 CC proliferation inhibiting nucleotide inserts in an example from the
 CC present invention. Methods from the present invention can be used to
 CC identify a proliferation- required gene in a microorganism, by contacting
 CC a microorganism with a proliferation-required gene activity inhibitory
 CC nucleic acid identified in another organism, and determining if
 CC inhibition occurs in the second microorganism. The nucleic acid sequences
 CC identified as being required for bacterial growth and proliferation, can
 CC be used for antisense therapy for killing bacteria.
 CC
 SQ Sequence 2383 AA;
 Query Match 4.5%; Score 118; DB 21; Length 2383;
 Best Local Similarity 19.5%; Pred. No. 1;
 Matches 95; Conservative 54; Mismatches 160; Indels 178; Gaps 19;
 QY 106 TSAAILISDPGQOPLAFNGLGS-----NTGL---QALG----- 136
 DB 1086 TLLATATVADPNHPAGITVFTWPDVAAANFTLNNNSIAITQANGAHTVLKGGKAGTH 1145
 QY 137 -ITDLDLGAQCPDQO---FADAANIAGDDLQSMVAFPTGTTIHGVLIGSDQDDFLDQFT 191
 DB 1146 TTVATLGNNTNSDSQPTVFVADKASQVNVLOISKDEITGNGVDSATLATATYKDOF-DHEV 1204
 QY 192 DDISSTFGSSITVOVALSGSARPPDQAGHEHFGFLDGISQPSVTGWETTVPFGQAVPPG 251
 DB 1205 NNLPVTFFSSASSGLTLPFGVSNNTN-----SGIAQATLAG---VAFGEKTVTAS 1250
 QY 252 IILRGDDDTGTRPSWALDGSFMAFRHFOQKVPFNATLANAIYANGAGNLTQEGAEF 311
 DB 1251 LANNAGS-----DNKTVHFTIGTAAAKI----- 1273
 QY 312 LGAMFGWKS GAPIDLAPADPALGADPQRNNNFVDSLTLDETRCPF----- 361

DB 1274 -----IELAPVDSIIAGT-PONSSGSVITATVDDNNCFPVKGVTVNFTSN 1318
 QY 362 GAHVKTNPRO---DLGCVDDTFHAMRSSIPYG--PETSDAELASG----- 402
 DB 1319 AATREMTNGQAVTNEGKATVITYTNTRSSIESGARDPVEASLENSSTLSISINVAD 1378
 QY 403 -----VTAODRGLLFVEYQSIIGCFRFOQIINNANNANPFSKPIPTGIE 447
 DB 1379 ASTAHLTLGLALPFTVSAGETTSLEYKDNXNGVPGQEVTLISVP-----SEGVTPSNN 1434
 QY 448 PIIQOTTPRTVGLDPIJNOMETFTVPLFVIPKGEY-----FPLPSISA 491
 DB 1435 AI--YTT-----NHGDNFYAS-FIATKAGVQLTATLENGDSMOOTVTVVPMVAN 1481
 QY 492 LTATIAA 498
 DB 1482 AETTLAA 1488
 RESULT 10
 ABE28940
 ID ABE28940 standard; Protein; 567 AA.
 XX
 AC ABE28940;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #28931.
 XX
 KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 XX
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Dzmanac RT, Liu C, Tang YT;
 PI
 XX WPI; 2001-639362/73.
 DR N-PSDB; AAS93127.
 XX
 PS New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20; SEQ ID NO 59299; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity

PF 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSEQ-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR N-PSDB; AAS87576.
XX
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX
PS Claim 20; SEQ ID No 53748; 103pp; English.
XX
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridization probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 1194 AA;

Query Match 4.3%; Score 111.5; DB 22; Length 1194;
Best Local Similarity 23.3%; Pred. No. 1.4;
Matches 84; Conservative 38; Mismatches 148; Indels 91; Gaps 17;

QY 91 TSFKTALKTVVPORTSAA-----ILISDP--SQQPLAFVNLGFSNTGLQALGTTD 139
Db 842 TTFQSLSHTAVPTATSSGAADFSRGSTLATSAPATSSQP---TLTFSNMTSTPTFNT-- 895
QY 140 DLGDAQFPDQGFADAAANLGDLSQVNAFTGTTTGHVFLIGSDDDFLDQFTDDISSTFG 199
Db 896 -----PFGSSAKSP-----LPSYPGANQPAF--GAAGQPPCAAKPALAPSF 937
QY 200 SSITQVALSGSARP-----FDQAGHEHFGF-----LDGI 229
Db 938 SSFTFGNSAAPAAAFTPAPSPMIKVPAVPTPIHPITFGATHSAFGKATASAFGAPAS 997
QY 230 SQPSVTGHETTVPQGAVVPGIILTGRDGTGTRPSWALDGSFMAFRHFOOKVPEFNAY 289
Db 998 SQPAF--GGSTAVFPG--AATSSGFGATQTASSGSSSS-----VFGSTTP--SPF 1042
QY 290 TL-ANAIPIANSAG---NLTOEGAEFLGARMGRKSGAPIDLAPTA---DDPALGADPQ 342
Db 1043 TFGGSAAPAGSGSGFINAVATPGSSTTTGAFSGAGSGSSTATSTPFAAGLGONALGTTGQ 1102
QY 343 RNNNEDYSDTLTDETRCPFGAHRKTNPRODLCGPVDTFHAMRSSIPYGPETSDAELASG 402
Db 1103 -STPFAFVSSSTTESKVPFGG---TATPTFGIANTPAFGVGTSGSSLSFGASSAPAGFVG 1158
QY 403 V 403

Db 1159 V 1159

Search completed: November 27, 2002, 13:47:07
Job time : 44 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Comugen Ltd.

OM protein - protein search, using sw model

Run on: November 27, 2002, 13:45:31 ; Search time 22 Seconds
(without alignments)
2176.134 Million cell updates/sec

Title: US-09-926-084-7
Perfect score: 2601
Sequence: 1 MDLSLFVSVAVLVGSSSHV.....KGEYFFLPISALTATIAA 498

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	412.5	15.9	460	2 D75610	probable peroxidase
2	386.5	14.9	469	2 AC2004	hypothetical prote
3	185	7.1	421	2 E82023	probable periplasm
4	177	6.8	421	2 F81244	conserved hypothe
5	143.5	5.5	434	2 AE0226	conserved hypothe
6	133.5	5.1	423	2 A99787	hypothetical prote
7	133.5	5.1	423	2 A85647	hypothetical prote
8	130.5	5.0	423	2 A64844	ycdb protein precu
9	127.5	4.9	1067	2 D75625	probable extracell
10	126.5	4.9	299	2 AH0367	conserved hypothe
11	121.5	4.7	1093	2 B86748	hypothetical prote
12	119.5	4.6	420	2 T50585	probable membrane
13	118.5	4.6	2468	2 A83412	hypothetical prote
14	118	4.5	2383	2 D64962	probable membrane
15	116	4.5	2660	2 E85822	hypothetical prote
16	115	4.4	3972	2 S75251	hypothetical prote
17	112.5	4.3	1248	2 B96837	hypothetical prote
18	112.5	4.3	1832	2 T31113	mucin-like glycopr
19	111.5	4.3	2712	2 T05113	hypothetical prote
20	110	4.2	4199	2 S76412	hypothetical prote
21	109.5	4.2	421	2 AC1481	conserved hypothe
22	109.5	4.2	421	2 AH1120	B. subtilis ywdn p
23	109.5	4.2	1742	2 S76110	hypothetical prote
24	108.5	4.2	3029	2 S76109	hypothetical prote
25	108	4.2	647	2 A45244	exo-alpha-sialidas
26	108	4.2	778	2 T17679	proline-rich prote
27	108	4.2	1345	2 H80975	hypothetical prote
28	107.5	4.1	2523	2 F70846	probable ppe prote
29	107	4.1	940	2 AD1374	internalin protein

30	107	4.1	3562	2 A47171	chondroitin sulfat
31	106	4.1	743	2 T34853	probable fusidic a
32	105.5	4.1	700	2 H85732	hypothetical prote
33	105.5	4.1	859	2 AE2217	hypothetical prote
34	105.5	4.1	2232	2 T34434	hypothetical prote
35	105	4.0	753	2 B85782	catalase, hydroper
36	105	4.0	753	2 F90933	catalase HPII limp
37	105	4.0	835	2 JC6140	cell surface-assoc
38	105	4.0	1965	2 S75200	fat protein - syne
39	104.5	4.0	409	2 E83179	hypothetical prote
40	104.5	4.0	430	2 S52700	NUF4 protein - ye
41	104.5	4.0	1684	2 S10789	amylase A-180 - al
42	104	4.0	603	2 S27055	basic serine prote
43	104	4.0	658	2 H84385	dipeptide ABC tran
44	104	4.0	753	2 A39129	catalase (EC 1.11.
45	104	4.0	2554	2 AB3528	extracellular seri

ALIGNMENTS

RESULT 1
D75610
probable peroxidase - Deinococcus radiodurans (strain R1)
C/Species: Deinococcus radiodurans
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Jun-2000
C/Accession: D75610
R/White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; S.; Shen, M.; Yamanevan, J.U.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; May, S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A/Title: Genome sequence of the radioreistant bacterium Deinococcus radiodurans R1.
A/Reference number: A75250; MUID:20036896; PMID:10567266
A/Accession: D75610
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-460 <WHI>
A/Cross-references: GB:AE001862; GB:AE001825; NID:g6460468; PIDN:AAF12218.1; PID:g646051.
A/Experimental source: strain R1
C/Genetics:
A/Map position: 2
A/Map position: 2
C/Superfamily: Deinococcus radiodurans probable peroxidase DRA0145

Query Match 15.9%; Score 412.5; DB 2; Length 460;
Best Local Similarity 29.3%; Pred. No. 1.1e-23;
Matches 148; Conservative 68; Mismatches 190; Indels 99; Gaps 25;

QY 34 PLTNFPGQAPLPTLQHTTESGANDTI-LPLNNIGDILVGMKKKER-FVFOVNDAT 91
Db 2 PEMTLF---KTLRELVIH-----NDKIDLDLDIQTATL-----RERPEPYGTAMV 46
QY 92 SFKTA-----LKTVPQRITSAAILISDPSQPLAFVNLGFSNTGLQALGITDLDGAQ 145
Db 47 RFDTAEGGRELKRLLP-HIASA-----EKMDVAVYAMTAAMISYEGKKLGVPODSDLS- 100
QY 146 FPD-----GQFADANLGD-----DLSQWVAPFTGTTIGVPLIGSDODD-----FLDQFDD 193
Db 101 FPEFVKYMGAGRAHLPDVGENDPKHWEKPGTGQVHALRTIFANEENMKALVIAHE 160
QY 194 ISSFFGSSITOVQALSSGARPFDOAGHEHFGFLDGISQPSVYGETTVFPGQA-VVPGGI 252
Db 161 LEATKG-VTILMEDDGAQP---DSRSLQYKDMISNPAIEGSGIRFPQGAIRKGE 215
QY 253 ILTRGDGDTG-----RDS-WALDGSFMAFRHFOOKVPEFNAVYLANAIPANSAGNLTQO 306
Db 216 FVLGPGAGVPLGMPKREVLGKNGTFVALRKYHTNAGSFYRLKENA--EYTGSD---- 269
QY 307 EGAEFLGRMGKMSGAPIDLAPTADDPALGADPPORNNNDYSDTLTDFRCFGAAVR 366
Db 270 --AELLAALKVGRWRSQAPLTLAPKEDDPGLGHDPNNRNDFTYKND--PEGLEVPFGSHIR 326
QY 367 KTNPRO--DLGGVDFPFHAMRSSIPYGPTSDAELASGVTAQRGILLFVEYSIIINGGF 423

Db 327 RMNPRTKLELLTDVNIHRIIRATAYGAYDPKASDLAEDKVERGLYFIFISAKAMDTT 386
QY 424 RFOQINWANNANFPFKPIPTGPIEGTTPRTVGLDPL-----NONETFTVTP----- 473
Db 387 EFLQKEWINKANF-----IGQGSER-----DPIVGLQDEDLTFTLPKEPVRQ 428
QY 474 -----LFVLPKGGYFFFLPSISAL 492
Db 429 RLRCMDTFNVLRGEXYLFMPLSALS 453
RESULT 2
AC2004
hypothetical protein alr1585 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AC2004
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
N.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana-
baena sp. strain PCC 7120
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2004
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-469 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA077951.1; PID:g17135405; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr1585

Query Match 14.9%; Score 386.5; DB 2; Length 469;
Best Local Similarity 29.9%; Pred. No. 1.1e-21;
Matches 144; Conservative 57; Mismatches 185; Indels 95; Gaps 23;
QY 64 LNNIQGILVGMKKQKRVFV--FQVNDATSFKTALKTVYVQBITSA-----AILISDP 115
Db 28 LNDLOGNLLKGHRDHSVHLFLOKPEQVEVVKQWISQATQYITSAKKQADEAFKYRQK 87
QY 116 SQQLAFVNLGFSNTGLQALGIT--DDLGAQFPDGFAD--AANLGD-DLSQWVAPFTG 170
Db 88 GVSQDVANFPLSRHGVYLEIEFPQIPGDKPPRMGMKNEIRSSLODPKXIATWELGFQ- 146
QY 171 TTTHGVFLGSDQDDFLDQPTDDISSTFGSSITQVQALSGSARFP---DQAGH--EHFGF 225
Db 147 SEIHALVLIADDDIVDLQIVNQIT---QKLQIAIEIVHREDGFIILNQAGQIIEHFGF 202
QY 226 LGDISQPSVTGWETTVPQGVV-----PPGIIILTRDQDGTGRPSWAL 269
Db 203 VDGVSQP-----LFMKRDVVYRVRVNNCDFDKWDPKAPLDSILV-EDPNGNTKDSY-- 251
QY 270 DGSFMAFRHFOOKVPSEFNAYTLANAI PANSAGNLTQEGAEFLGARMFGKWSGAPIDLA 329
Db 252 -GSYLVRKLEQNVKAFREDQKLAOKLNLQENLA-----GALIVGRFADGTPTVL- 301
QY 330 PTADDDPALGADPORNNDYSDTLTDETRCPFGAHVRKTNPRQOLGGPVDT----- 380
Db 302 --SDIPTYATP--TNNFNVDGLA-ATKCPFHSHTRKTNPRGDTARLLTTDGHFDEAPK 356
QY 381 -----FIAMRSSIPYGPETSDAELASGVTAQDRGLLFVEYQSIIGNFRFOQINWANNANF 436
Db 357 EERGHRTIRAVSYGENNPSKEPVSG-----SGLLFLCFQSNFENFNMQSRWANPQNF 411
QY 437 PFSKPIPTGPIEPIIGT-----PRTVGGLDPLNQNETFTVPLFVPKGGYFFFLPSISA 491
Db 412 ---VQVNTGDPDLIGPSGTQKPKKWG---EP---ETEENFQWLNWNNKGGYFFAPSISF 464
QY 492 L 492
Db 465 L 465

RESULT 3
EB2023
probable periplasmic protein NMA0282 [imported] - Neisseria meningitidis (strain Z2491)
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: EB2023
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel,
P.; Holroyd, S.; Jagers, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
A.; et al. Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: AB1775; MUID:20222556; PMID:10761919
A:Accession: EB2023
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-421 <PAR>
A:Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAB83589.1; PID:g7379044
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: NMA0282
C:Superfamily: hypothetical protein ycdB

Query Match 7.1%; Score 185; DB 2; Length 421;
Best Local Similarity 23.2%; Pred. No. 2.2e-06;
Matches 88; Conservative 41; Mismatches 125; Indels 126; Gaps 17;
QY 135 LGITDDLGAQFPDGFQFA-----DAAN-----LGGDLSQWVAPFTTTIH 174
Db 130 LTVTVGSGSLF-DGRFGLKDKKPIHLQEMRDFSNLQKQSCDGLSLQICAFPTETCQ 188
QY 175 GVFLGSDQDDFLDQPTDDISSTFGSSITQ--VQALSGSARFPDQAGHEHFGFLDISQP 232
Db 189 AA-----LRDIIKHTVQTAVIRWSIDGQPKSEFGMAARNLLGFRDGTGNP 235
QY 233 SVTGWETTVPQGVVPPGIIILTRDQDGTGRPSWALDGSFMAFRHFOOKVPEFNAYTLA 292
Db 236 KVSQDKTA-----DEVLTGVAANSLDEPEWAKNGSYQAVLRIRHFVFWDTPL- 285
QY 293 NAIPANSAGNLTQEGAEFLGARMFGKWSGAPIDLA PTDADDPALGADPORNNDYSDT 352
Db 286 -----QEQTDFGRRKY---SGAPMDGKKEADQDFAKDPEGNTT----- 322
QY 353 LTDETRCPFGAHVRKTNPRQDLGGP---VDTFHAMRSSIPYGPETSDAELASGVTAQDRGL 410
Db 323 -----PKDSHIRLANPRD-----PEFLKKHLFRAYSY-----SRGLASSG--QLDVG 365
QY 411 LFVEYQSIIGNFRFOQINWANNANPPFSKPIPTGPIEPIIGTTPRTVGGLOPLNQNETF 470
Db 366 VFCVQANLADGFIFVQ-NLLNG-----YISPFGGYFFFLPGV 408
QY 471 TVPLFVLPKGGYFF-LPSI 489
Db 393 ---YISPFGGYFFFLPGV 408

RESULT 4
F81244
conserved hypothetical protein NMB0036 [imported] - Neisseria meningitidis (strain MC58)
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: F81244
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.;
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
Ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masigiani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; V.
et al. Nature 404, 502-506, 2000
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: AB1000; MUID:20175755; PMID:10710307
A:Accession: F81244
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-421 <TET>
A:Cross-references: GB:AE002364; GB:AE002098; NID:g7225269; PIDN:AAF40507.1; PID:g72252.

C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C/Accession: A85647

R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001

A/Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.

A/Reference number: A85480; MUID:21074935; PMID:11206551

A/Cross-references: GB:AE005174; NID:g12514381; PIDN:AAG55637.1; GSPDB:GNO0145; UNGP:Z15
A/Experimental source: strain O157:H7, substrain EDL933

C/Genetics:

A:Gene: ycdB

C/Superfamily: hypothetical protein ycdB

Query Match 5.1%, Score 133.5; DB 2; Length 423;
Best Local Similarity 22.3%; Pred. No. 0.018;
Matches 81; Conservative 29; Mismatches 119; Indels 135; Gaps 14;

Qy 145 QPDPQFADANLGGDLDSQWVAPFTGTTIHGVFLIGSQDDFL-----DOFTDDISSTF 198
Db 162 RFPNDSLDAALCHGDVLLQICANTQDTVIHALRDIHKHTPDLLSVRWKREGFISDHAARS 221
Qy 199 GSSITQVALSGSARPFDOAGHEHFGFLDGLGISQPSVTG-----WETTVPQAVVPP 250
Db 222 KGKETPINLL-----GFKDGTANPDSQNDKLMQKVWVTA----- 256
Qy 251 GIILTRDGDGTTPRSWALDGSFNAHFQOKVPEFNAYTLANAIPANSAGNLTOQSGAE 310
Db 257 -----DQEPAMWTIGGSQAVRLIQPRV-EFWDRT-----PLAEQQ-- 291
Qy 311 FLGARMGRWK-SCAPIDLATADDPALGADPQRNNFNFDYSDTLTDETRCPFGAHVRKTN 369
Db 292 ----TIFGRDKQTGAPLGMQHEHDVPVDSPEKG-----TALDSHIRLAN 334
Qy 370 PRDLGGPVDTTFHAMRSSIPYGPETSDAELASGVTAQ----DRGLLFVEYQSIINGGRFFQ 426
Db 335 PRT---AESESSLMLRGYSY-----SLGVTNSGOLDMGLLFVYQHDLEKGLTV 382
Qy 427 QINWANNANFPSPKPIITGIEPIIQCTPTRVGGDLPLNQNETTVPFLVPIPKGGEYFF- 485
Db 383 QKRLNGEALSEYVKPI-----GGYFFA 405
Qy 486 LPSI 489
Db 406 LPGV 409

RESULT 8

A64844

ycdB protein precursor - *Escherichia coli* (strain K-12)

C/Species: *Escherichia coli*

C/Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002

C/Accession: A64844; A47065

R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997

A/Title: The complete genome sequence of *Escherichia coli* K-12.

A/Reference number: A64720; MUID:97426617; PMID:9278503

A/Accession: A64844

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-423 <BLAT>

A/Cross-references: GB:AE000203; GB:U00096; NID:g1787248; PIDN:AAC74104.1; PID:g1787255
A/Experimental source: strain K-12, substrain MG1655

R/Kim, S.K.; Makino, K.; Anemura, M.; Shinagawa, H.; Nakata, A.
J. Bacteriol. 175, 1316-1324, 1993

A/Title: Molecular analysis of the *phoH* gene, belonging to the phosphate regulon in *Esch*

A/Reference number: A47065; MUID:93186698; PMID:8444794

A/Accession: A47065

A/Status: preliminary

A:Molecule type: nucleic acid
A:Residues: 196-423 <KIM>
A:Cross-references: GB:D10391; GB:D90448; NID:g285772; PIDN:BAA01229.1.; PID:g285773
A:Notes: sequence extracted from NCBI backbone (NCBIN:126684, NCBP:126688)
C:Genetics:
A:Gene: ycdB
C:Superfamily: hypothetical protein ycdB
F:1-35/Domain: signal sequence #status predicted <SIG>
P:36-423/Product: ycdB protein #status predicted <MAT>

Query Match 5.0%; Score 130.5; DB 2; Length 423;
Best Local Similarity 22.3%; Pred.No.0.031;
Matches 81; Conservative 29; Mismatches 119; Indels 135; Gaps 14;

QY 145 QFPDQGADAAANLGDLSQWAPFTGTTIHGVFLIGSDQDDPL-----DOFTDDISSTF 198
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
Db 162 RFPNDSLDAALCHGDVLLQICANTODTVIHALRDIHKHTPDLLSVRWKREGFISDHARS 221
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
QY 199 GSSITQVALSGSARPPDQAGHEHFGFDGISQPVSVTG-----WETVPFGQAVVP 250
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
Db 222 KGKETPINL-----GFKDGTANPDSONDKLMQKVVMVTA----- 256
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
QY 251 GIILTRDGDTGRPSHWALDSFMAFRFQOKVFEPFNAYTLANAIPANSAGNLTOEGAE 310
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
Db 257 -----DQQEFAMTIGGSQAURLIQFRV-EFWDR-----ELKEQO---- 291
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
QY 311 FLGARMFGRWK-SGAPIDLATDPDPAAGDPORNNNFYSDTLTDCTCPFGAHVRKTN 369
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
Db 292 ----TIQRDKQTGAPlOMQHEDVDVASPECK-----VIADSHIRLAN 334
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
QY 370 PROLDGGPVDTTHAMRSSIPYPETSDAELASGVTAQ---DRGLLFVEYSIQINGFRFO 426
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
Db 335 PRT--AESSESLMLRCGY------SLGVTNSQLDMGLLVFCVQHDLEKGFLTIV 382
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
QY 427 QINWANNANPFSPKIFPGICEPIIGTCOTPRTVGGDLPLNQNETPTVPVLFIGGEYFF- 485
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
Db 383 QRUNGEALEEYVKPI-----GGGYFFFA 405
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
QY 486 LPSI 489
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
Db 406 LPGV 409

RESULT 9
D75625
probable extracellular nuclease - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: D75625
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zaleski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A>Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; UID:20036896; PMID:10567266
A:Accession: D75625
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1..1067 <WHI>
A:CROSS-references: GB:AEO01826; NID:g6460827; PIDN:AAF12592.1; PID:g6460889; TIGR:DRB8
A:Experimental source: strain R1
C:Genetics:
A:Gene: DRB0067
A:Map position: megaplasmid
A:Genome: plasmid
A>Note: plasmid MP1

Query Match 4.9%; Score 127.5; DB 2; Length 1067;
Best Local Similarity 23.0%; Pred.No.0.2;
Matches 121; Conservative 60; Mismatches 171; Indels 175; Gaps 28;

QY 13 LVGSSSHVNAA---KLGARQTRFTPL-----LTNPFGQAP 44
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :


```

Db 55 LSGSAIKKASVRLQAQATEVGLSKSNNGYVDDSNNTLRMMATPVTYTSDD-P 113
Qy 45 L--PTLQHTTESSANDTI--LPLNNIOGDLIVGKKOKERFVFOVNDATSF----- 93
Db 114 LAATPFI-PVDE -GAGGIGTTAFKVR-----YFDSDASSRRLPLAFD 157
Qy 94 ----KTKLKYVPQRITSAAIILSDPSQOPLAFVNLGFSNTG--LQAL--GITDDLGA 144
Db 158 SSNDPKTAVKTKQLLRDLSDSGVQVNLPAQLQLA---GTSHTGMWGMALPAGAT----- 207
Qy 145 QFPFGQADANLDDLSQWAPFTGTTIHGVFLIGSDQDDPFLQDFDIDISTGSSITQ 204
Db 208 ----GQVTLAASIPASISQDNPPS-----FDLVFTAVANVSTLTNTGA 249
Qy 205 VQALS--GSARPFDOAGHEHFGFLDGISQSVYGTWETTFPGQAVVPGLITLG---RDG 259
Db 250 VQCGTTPGGDRRAALTGAG--TIEGVTAIVERG-----LSGFFVQDEG 289
Qy 260 DTGTRPSMALDGSFMAFRHFOQKVPFNAYTLNAILPANSAGNLTQEG--AEFLGARMF 317
Db 290 IDADRDETTSGLFV-----YCAASC-PALISAGDRVRVSGTVAEYGGATQM 334
Qy 318 ----GRKSGAPIDLAFTADDPALGADPQNNND-----YSDTLDETSCPGAHYRK 367
Db 335 TAPVTYKLSGLAL--PPAAELKLPDKTQOERYEGMRKVPFETLTITNNYTYGRY--- 388
Qy 368 TNPQDLDGAPVDTFHAMRSI-PYGERTSDA-----LASGTAQ--DRGLLFVEYQSI 419
Db 389 -----GQDLDSNAGMFPNTGNASASQSTITLDDGVSAGNPLDLNLSERTRT 440
Qy 420 GNGFRFOQIMNANNAPFSPKPTPGIEPIIGOTT-----PRTVGG 460
Db 441 GDVTYGLSGVMHVSANQMLEP--EGAVEFVSANSRASNALPHVDGG 485

```

```

RESULT 10
AH0367
conserved hypothetical protein YP03025 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C:Accession: AH0367
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
demo-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Status: preliminary
A:Accession: AH0367
A:Molecule type: DNA
A:Residues: 1-299 <KIR>
A:Cross-references: GB:AL590842; PIDN:CAC92267.1; PID:G15980978; GSPDB:GN00175
A:Gene: YP03025
C:Superfamily: Escherichia coli hypothetical protein b2431

```

```

Query March 4.9%; Score 126.5; DB 2; Length 299;
Best Local Similarity 21.7%; Pred. No. 0.038;
Matches 91; Conservative 45; Mismatches 109; Indels 175; Gaps 23;
Qy 109 AILSDPSQOPLAFVNLGFSN--TGLOALGITDDLDGAQFPD---GQFADANLGDLS 162
Db 16 AIFMAKVOGELDAIRLCKCKRCQSLQEL-----QQQFPBHGAVYARGSNWHLS 68
Qy 163 ----QWAPFTGTTIHGVFLIGSDQDDFL--DQFTDISSTFSSITVOVALSGARP 214
Db 69 NGQGAKEKLPFVPL--GKGLAPATQRDLILHISLRQDINFLLQAQAV--AAGSAIA 122
Qy 215 FDQAGH-----EHFGLDISQPSVTGMETTFVPGQAVVPFGIILTRGCDGTSPS 266
Db 123 VEEETHGRFWRVEERDFTGIDTENP-----QGDKRPEVAIADG-----EEDAG--- 167
Qy 267 WALDGSFMAFRHFO-----QKVPFNAYTLNAILPANSAGNLTQOEGAEFLGAMFGRM 320

```

```

Db 168 ----GSYLVQREYEHNLNKKQRIPE-----NEDE-----KIIGRT 198
Qy 321 KSGAPIDLAFTADDPALGADPQRNNNPFYDSYDLTDETRCPFGAHVRTNRODLCGPVT 380
Db 199 K-----LBSQELPSD-OR-----PDTSHVSRVDLKENKG----- 227
Qy 381 FHAMRSIPYGPETSDAELASGVTAODRGLLFVEY-----QSIINGFRFOQIMAN 432
Db 228 LKILRQSLPYG-----LASG-----KHGLYFIAYCARLHNEQQLLS---MFGDIDGKH 273
Qy 433 NANPFSKPTTPGIEPIIGOTTPTTVGGLDPLNQNFTVPLFVIPKGEYFPLPSIAL 492
Db 274 DQLRFRSKPVT-----GSYFAPSLTAL 296

```

```

RESULT 11
B86748
hypothetical protein ykbc [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C>Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: B86748
R:Bioclin, A.; Mincker, P.; Mauger, S.; Jallion, O.; Malarme, K.; Weisenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssi
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: B86748
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1093 <STO>
A:Cross-references: GB:AE005176; PID:G12723930; PIDN:AAK05084.1; GSPDB:GN00146
A:Experimental source: strain IL1403
A:Gene: ykbc

```

```

Query March 4.7%; Score 121.5; DB 2; Length 1093;
Best Local Similarity 21.3%; Pred. No. 0.59;
Matches 123; Conservative 64; Mismatches 228; Indels 163; Gaps 27;
Qy 17 SSHNAKLAGRCQRTTP-LITNFGQAPPLTLTO-HTTESGANDTILPLNNIOGDLVG 74
Db 62 SPTVNSSLAASDSTTEBPAAVSNMENTAOUSSSTNLTSTSTVDITNSPTQSNQNSIAIS 121
Qy 75 MKQKERRFVFOVNDATSPKTAIKTYVPORITSAAIILSDPSQOPLAFVNLGFSN----- 129
Db 122 LTKASRESSINERPASTIANETTTADVTQTITAQAQTSPTTS-LSSSESGKRTNSQS 180
Qy 130 ----TGLOALGITDD--LGDAPPDGQFADANLGDLSQWAPFTGTTIHGVFLIGSD 182
Db 181 KSTEITNIGVAVDNNATGSAVF-----DGVNI-----TLQG----- 213
Qy 183 QDPFLDQFTDISSTFGSSITVOVALSGSARFPQAGHEHFGFLDGISQPSVTGMETTFV 242
Db 214 -KDITDDLDLPSGLHWSQCTQVIAIKGTAT--GQLNQENFGVQDGPILPAT--YTWN 267
Qy 243 PGQA--VVPFGIILTRGPD-----TGTRPSW--ALDGSFMAFRHFOQKVPFNAYTLAN 293
Db 268 DQDSGRITVYKKSLSGLDLDMITVVASDQSWQANBCA-----EGIQGLTFGEQ 319
Qy 294 AIPANSAGN--LTOEGAEFLGAMFGRWKSAPIDLAFTAD-----DPALGAD----- 340
Db 320 NI--AESGNSIVCLYNGMNL-SLIYQVGHDTTTEVPVVASFITTDIDNAGVQTNLAN 377
Qy 341 ----PQRNNPFYSDTLTDET-----RCPFGAHV----- 366
Db 378 LVTLIPQTLNKKQDGTIYDASPNVPGLDVASLPYGGVYLCAGFVSEFYVNVAPAPER 437
Qy 367 ----KTNPQDLDGAPVDTFHAMRSI-----PYGERTSDAELASGVTAODRGL 410
Db 438 DQSFPAQGVRYDLPGSALQAH--MNTQIRQNFYVYDFEFHKLQETDHYLGFTGQDYNL 496
Qy 411 LFVEYQSIINGNF-----RFOQIMNANNAP-----FSKPTTP 444

```

Db 497 ---PIPTIKGFGFVNLTENDASKNNPVINLIYNHNLPTYGNHNNNIYYQGTAYTPSFTI 553

445 GIEPIIGOTTPTRTVCGGLDPLNQNETFTVPLFVIPKGG 482 QY

D**b** 554 GYQNI--GNPEASITYTPVNNKGASSVTLPIMAGK 588

RESULT 12

T50585
 probable membrane protein [imported] - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 28-Jul-2000
 C:Accession: T50585
 F:Redenbach, M.; Kieser, H.M.; Denapaita, D.; Eichner, A.; Cullum, J.; Kinashi, H.; Hopwood
 Mol. Microbiol. 21, 77-96, 1996
 A:Title: A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb S
 A:Reference number: Z20556; MUID:97000351; PMID:8843436
 A:Accession: T50585
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-420 <RED>
 A:Cross-references: EMBL:AL133220; PTDN:CAB61722.1
 A:Experimental source: strain A3(2)
 C:Genetics:

RESULT 13
A83412 hypothetical protein PA1874 [imported] - Pseudomonas aeruginosa (strain PA01)
C/Species: Pseudomonas aeruginosa
C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C/Accession: A83412
R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bu
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbige, K.; Lim,
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A/Reference number: A82950; MUID:20437337; PMID:10984043
A/Accession: A83412

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2468 <STO>
A:Cross-references: GB:AE004613; GB:AE004091; NID:G9947856; PIDN:AAG05263.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1874

Query Match	4.6%	Score 118.5	DB 2	Length 2468
Best Local Similarity	21.1%	Pred. No. 3.2		
Matches 115	Conservative 57	Mismatches 218	Indels 155	Gaps 28
Qy	21	NAAKL-GAROTRTPLLTNFPQGAPLPTLQHTTESGANDTILPLNNIQGDILVGMKKOK	79	
Db	1088	NGSSLSTAPGSGTVILTDNGN-----PIAEVTADSGNWYTPSTPIANGTVNVVAQ-	1142	
Qy	80	ERPVFFVQNDATSEFKTALKTYVPQRITSAAILSDPSQQPLAFVN--LGFNNTGLOALGI	137	
Db	1143	-----DASG-----NSSPPATVTVDSSAPPAPVNPNSGVVISGTAEGA	1182	
Qy	138	TDDLGDGA-QPDPGOFADAANLGGDLDSQWAPPT-GTTI-HCVFLIGSDQDDPLDQDTDDI	134	
Db	1183	TVLITDAGGNPIQGVTT-----ADSGNW--SFTPGTPLANGTVIVAT-----ATDPT	1227	
Qy	195	SSTGSSGITVQALSGSARPFDOAGHEHFGLDGISOPSVTGWETVFPQAVVPQIIL	254	
Db	1228	GNTGQQAATTVDAAVAPAPVID-----PS-----NGTISTGAEGAKVIL	1268	
Qy	255	TGRDGTGTRPSWALDGSFWAFRHFOQKVPFENAYTLANAI PANSAGNLTQQSGAEFL--	312	
Db	1269	T--DNGNPIGETTADGS--GNNSFTPGTPLANG-TVNVAQDPAGN-TGPGSGTTVDA	1322	
Qy	313	-----GARMFGRWKSAPIDLATDADPALGADPQRNNNFYSDTLTDETRCP	360	
Db	1323	VAPHTPVNPSGNLLNGTAEPSGTVTLTDGNGNPIQGTTADSGNWSF---TPGSQLP	1378	
Qy	361	FGAHRVKTNPQDLGG---PVDTFHAMRSSIPYGPET--SDAELASGVTAQDRGLLFVE	414	
Db	1379	NGTVNVVT--ASDAAGNTSLPAIT--TVDSLSLSIPQVDPNSGVSIGTADAGNTIIITD	1434	
Qy	415	YQSIIGNGFRPQOI-----NWANNANFPFS-----KP	441	
Db	1435	-----GNGNPIGQVTDADSGNWSFTPGIPLDGTVVNVVARSPSNVDSAPAVITVDGVAP	1489	
Qy	442	ITPGIEPI-----IGOTPTRTVGGDLPLNQNETFTVPLFVIPKGEYFELPSISA	491	
Db	1490	AAPVIDPSNGTEISGTAEGATVILTDGNGNPIGQ-----ATADSGSNWTFTTSTPL	1541	
Qy	492	LTATI	496	
Db	1542	ANGTV	1546	

RESULT 14
D64962
probable membrane protein b1798 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: D64962
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C.
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A67720; MUID:97426617; PMID:9278503
A:Accession: D64962
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Molecule type: DNA
A:Residues: 1-2383 <BLAT>
A:Cross-references: GB:AE000289; GB:U00096; NID:G1789285; PID:G1789288,
A:Experimental source: strain K-12, substrain MG1655
C:Keywords: nucleotide binding; P-loop; transmembrane protein
F:54-70/Domain: transmembrane #status predicted <TM>
F:1564-1571/Region: nucleotide-binding motif A (P-loop)

Query Match 4.5%; Score 118; DB 2; Length 2383;
 Best Local Similarity 19.5%; Pred. No. 3.3;
 Matches 95; Conservative 54; Mismatches 160; Indels 178; Gaps 19;

```

QY 106 TSAAILSDSPSOQPLAFVNLGFS-----NTGL---QALG-----136
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1086 TTTLATATKDPDSNHPVAGITVFTMPQDVANFTLENNGIAITQANGEAHTLKGKAGTH 1145

QY 137 -ITDDLGAQFPDQGFADANLGDLSQWVAFPTGTTIHGVFLIGSDQDDELDOFT 191
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1146 TVTATLGNNTSDSQPTFVADKASQVVLQISKDEITGNVDSATLTATVKQOF-DNEV 1204

QY 192 DISSTFGSSITVOVALSGSARPPDQAGHEHFGFLDGISQPSVTGWETTVPPGQAVVPPG 251
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1205 NNLPVFTSSASGLTLPFGVNTNE-----SGIAQATLAG---VAFGEKTVTAS 1250

QY 252 IILTRGDGTGTRPSWALDGSFMAFRHQKVPFENATTLANAIPANSAGNLTOEGAEF 311
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1251 LANNAS-----DNKTVHFIGDTAAKI-----1273

QY 312 LGAMFRMWSGAPIDLAPTDALGADPQRNNNFDVSDTLTDETRCF-----361
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1274 -----IELAPVPSIIAGT-PQSSGSVITATVDNNGFPVKGVTVNFTSN 1318

QY 362 GAHVKTNPQ---DLGSPVDTFHMRSSIPYG--PETSADALAG-----402
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1319 AATMEMNNGQAVTNEQKATVTVNTRSSIESGARPPTVASLENGSSTLSTSNVAD 1378

QY 403 -----VTAQDRGLLFVEYQSIIGNGFRFOQINMANNANFPFSKPT 447
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1379 ASTAHLTLQALPDTVASGETTSLYEVDKNGVGPQOEVTLSVP---SEGVTSNN 1434

QY 448 PIIGQTPRTVGGLDPLNQNFTFVPLFVIRKGEY-----FLPISIA 491
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1435 AI--YTT-----NHDGNFYAS-FTATKAGVYQLTATLENGDSMQQVTVVNVAN 1481

QY 492 LTATIAA 498
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1482 AETILAA 1488

```

RESULT 15
 E85822
 probable_invasin_23135 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
 C/Species: Escherichia coli
 C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C/Accession: E85822
 R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousie, K.; Apodaca,
 Nature 409, 529-533, 2001
 A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A/Reference number: A85480; MUID:21074935; PMID:11206551
 A/Accession: E85822
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 12660 <STO>
 A/Cross-references: GB:AE005174; NID:912516151; PIN:AA057041.1; GSPDB:GN00145; UMG:Z31
 A/Experimental source: strain O157:H7, substrain EDL933
 C/Genetics:
 A/Gene: 23135

Query Match 4.5%; Score 116; DB 2; Length 2660;
 Best Local Similarity 18.2%; Pred. No. 5.5;
 Matches 114; Conservative 77; Mismatches 211; Indels 224; Gaps 26;

```

QY 11 AVLGSSSHVNAAKLGARQRTTPLL-----TNPQGAPLPTLTQHTT 54
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1245 AVVLQTSKAILIINGVETITLATVCKPFDNAVVDLQVTESTN-PADTQLSQSKSTND 1303

QY 55 SGAND-----TLPNNIOGILVGMKKQKRFVFFQVNDATSPFKALKYVPQRITS-- 107
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1304 SGVAEVPFKGTIVLGVTAEATLPNGNNDTK---IVNIAPDASNAQVTLINIPAOQVVTNNS 1360

```

```

QY 108 -----AAILSDSPSOQPLAFVNLGFS-----NTGL---QALG-----136
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1361 DSVQLTATVADPSNHPVAGITVFTMPQDVANFTLENNGIAITQANGEAHTLKGKAG 1420

QY 137 -ITDDLGAQFPDQGFADANLGDLSQWVAFPTGTTIHGVFLIGSDQDDELDOFT 187
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1421 THVTATLNNNTSDSQ--PVTFVADKTSALVVLQISKNEITGNVDSATLTATVKQOF- 1477

QY 188 DQFDDISSTFGSSITVOVALSGSARPPDQAGHEHFGFLDGISQPSVTGWETTVPPGQAV 247
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1478 DNEVNNLPVFTSSASGLTLPFGVNTNE-----SGIAQATLAG-----1516

QY 248 VPPGIIILTRGDGTGTRPSWALDGSFMAFRHQKVPFENATTLANAIPANSAGNLTOE 307
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1517 -----VAFGEQVTVASLANGA---SDNK 1537

QY 308 GAETLGARMGKWSGAPIDLAPTDALGADPQRNNNFDVSDTLTDETRCF-----361
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1538 TVHFIGDT---AAKIIELTPVPSIIAGT-PQSSGSVITATVDNNGFPVKGVTVN 1591

QY 362 ---GAHVKTNPQ---DLGSPVDTFHMRSSIPYG--PETSADALAG-----402
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1592 FTSNAATAEMNNGQAVTNEQKATVTVNTRSSIESGARPPTVASLENGSSTLSTSN 1651

QY 403 -----VTAQDRGLLFVEYQSIIGNGFRFOQINMANNANFPFSKPT 443
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1652 VNADASTAHLTLQALPDTVASGETTTLYEVDKNGVGPQOEVTLSVP---SEGV 1707

QY 444 PGIETP-----IGQTPRTVGGLDPLNQNFTFVPLF-----475
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1708 PSNAIYTTNHDGNFYASFTAXKAGVYQV-ATLENDXSMQQVTVY-VPVANAELSLAA 1765

QY 476 ----VIRKGEYFFLPISALATIA 497
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1766 SKDPVIANNND-----LTTLTATVA 1785

```

Search completed: November 27, 2002, 13:48:18
 Job time : 29 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 27, 2002, 13:45:31 ; Search time 14 Seconds
(without alignments)
1475.372 Million cell updates/sec

Title: US-09-926-084-7

Perfect score: 2601
Sequence: 1 MDLSLFVSVAVLVGSSSHV.....KGGYFPLPSIALTATIAA 498

Scoring table: BLOSUM62
Gap 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	130.5	5.0	423	YCDB_ECOLI	P1545 escherichia
2	118	4.5	2358	YEEB_ECOLI	P76347 escherichia
3	114	4.4	2660	YEEB_ECOLI	Q8X8V7 escherichia
4	108.5	4.2	953	LKX3_PASHA	P55116 pasteurella
5	108	4.2	647	NANH_MICVI	Q02834 micromonospora
6	107	4.1	3562	PGCV_CHICK	Q00953 gallus galli
7	105	4.0	1229	NI21_HUMAN	Q9Y2N3 homo sapien
8	104.5	4.0	430	NU42_YEAST	P49686 saccharomyc
9	104	4.0	603	BPRV_BACNO	P42779 bacteroides
10	104	4.0	753	CATE_ECOLI	P1179 escherichia
11	103.5	4.0	710	FOXA_YEREN	O01674 yersinia en
12	103	4.0	1076	NUPI_YEAST	P20676 saccharomyc
13	102.5	3.9	1015	FOU_HVIR	P03367 human immun
14	102.5	3.9	1036	HP12_DEIRA	P3126 deinoxococcus
15	101.5	3.9	1480	PANI_YEAST	P32521 saccharomyc
16	101	3.9	252	CLCD_RHOOP	O67988 rhodococcus
17	101	3.9	681	YADD_SCHPO	O09838 schistosom
18	101	3.9	890	SYA_DEIRA	Q9R827 deinoxococcus
19	101	3.9	1244	SLAI_YEAST	P32790 saccharomyc
20	100.5	3.9	1856	MGA_HUMAN	O43451 homo sapien
21	100	3.8	600	IF2P_HALNI	Q04874 halobacteri
22	100	3.8	1002	P04587 human immun	P76115 escherichia
23	99.5	3.8	700	YNGD_ECOLI	P36712 human adeno
24	99	3.8	582	HEX3_ADE12	P03366 human immun
25	98.5	3.8	1003	FOU_HVIR	P11079 reovirus (t
26	98.5	3.8	1289	MCE_REOVD	P04585 human immun
27	97.5	3.7	1003	PGCA_HVIR	P07897 reovirus norv
28	97.5	3.7	2124	PGCA_RAT	P18802 human immun
29	96	3.7	1002	POL_HVIR	P20892 human immun
30	95.5	3.7	1003	POL_HVIR	P20892 human immun
31	95	3.7	937	NU98_HUMAN	P52948 homo sapien
32	95	3.7	1003	POL_HVIR	P12497 human immun
33	95	3.7	2003	YDBA_ECOLI	P33666 escherichia

34	94.5	3.6	735	1	DH84_MOUSE	P51660 mus musculu
35	94.5	3.6	937	1	NU98_RAT	P49793 rattus norv
36	94.5	3.6	2090	1	N214_HUMAN	P35658 homo sapien
37	94	3.6	1002	1	POL_HVIR	P05959 human immun
38	94	3.6	1089	1	MMLE_MYCTU	O07800 mycobacteri
39	93	3.6	1002	1	POL_HVIR	P12499 human immun
40	93	3.6	1189	1	ITRA_HUMAN	Q9UXK5 homo sapien
41	92.5	3.6	959	1	N100_YEAST	O02629 saccharomyc
42	92.5	3.6	978	1	PM14_CHLPP	O92895 chlamydia p
43	92.5	3.6	995	1	Y109_YEAST	P40442 saccharomyc
44	92.5	3.6	1006	1	POL_HVIR	P05961 human immun
45	92.5	3.6	1015	1	POL_HVIR	P04587 human immun

ALIGNMENTS

```

RESULT 1
ID YCDB_ECOLI STANDARD; PRT; 423 AA.
AC P1545; P75903;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ycdB precursor.
GN YCDB OR B1019.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=9746617; PubMed=9278503;
RA Blatner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada H., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 7.18-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12,728,0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [3]
RP SEQUENCE OF 196-423 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=9316698; PubMed=844794;
RA Kim S.-K., Makino K., Amemura M., Shinagawa H., Nakata A.;
RT "Molecular analysis of the phoH gene, belonging to the phosphate
RL J. Bacteriol. 175:1316-1324(1993).
CC -1- SIMILARITY: TO B.SUBTILIS TWBN (IPA-29D).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000203; AAC74104.1; -.
CC DDB; D90738; BAA35796.1; -.

```



```

QY 403 -----VTAODRGLLFVEYOSIIGNGFRFOQIMANNANPFPSKPTTGCIE 447
DQ 1354 ASTAHLTLQLPDTVSAGERTSLYIEVKDNYNGVPGQEVTLASVP-----SEGVTPSNN 1409
QY 448 PIIQOTTPRTVGGDLPLNQNFTFVPLFVIRKGBEY-----PFLPSISA 491
DQ 1410 AI--YTT-----NHDGNFYAS--FTATKAGVYQLATLLENGDSMOQTVTYVPNVAN 1456
QY 492 LTATIAA 498
DQ 1457 AETTLAA 1463

RESULT 3
YEST_ECO57
ID_YEST_ECO57 STANDARD; PRT; 2660 AA.
AC Q8X8V7; Q8X2C0; Q8X2B9;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein yestJ.
GN 22135 OR EC62775/EC62776.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glaesner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Prost J., Hackert J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobbeck E.V., Davis N.W., Lim A., Dimalanta E.T., Petalousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki G., Ogasawara N., Yasunaga T.,
RA Kunara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -1 SIMILARITY: CONTAINS 16 BIG-1 DOMAINS.
CC -1 SIMILARITY: BELONGS TO THE INTIMIN/INVASIN FAMILY.
CC -1 CAUTION: Ref.2 sequence differs from that shown due to a
CC frameshift in position 1315.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
CC
DR EMBL; AE005423; AAG57041.1; -
DR EMBL; AP002559; BAB36198.1; ALT_FRAME.
DR EMBL; AP002559; BAB36199.1; ALT_FRAME.
DR InterPro; IPR003344; BIG_1.
DR InterPro; IPR000601; PKD_domain.
KM Hypothetical protein; Repeat; Complete proteome.
FT DOMAIN 738 834 BIG-1.1.
FT DOMAIN 840 929 BIG-1.2.
FT DOMAIN 931 1033 BIG-1.3.
FT DOMAIN 1042 1132 BIG-1.4.

```

```

FT DOMAIN 1134 1236 BIG-1.5.
FT DOMAIN 1245 1335 BIG-1.6.
FT DOMAIN 1337 1439 BIG-1.7.
FT DOMAIN 1448 1539 BIG-1.8.
FT DOMAIN 1548 1632 BIG-1.9.
FT DOMAIN 1653 1750 BIG-1.10.
FT DOMAIN 1751 1855 BIG-1.11.
FT DOMAIN 1856 1957 BIG-1.12.
FT DOMAIN 1963 2056 BIG-1.13.
FT DOMAIN 2065 2156 BIG-1.14.
FT DOMAIN 2157 2252 BIG-1.15.
FT DOMAIN 2254 2355 BIG-1.16.
SQ SEQUENCE 2660 AA; 280062 MW; 01EB92A08F5C09D2 CRC64;

Query Match 4.4%; Score 114; DB 1; Length 2660;
Best Local Similarity 18.5%; Pred. No. 4.1; Indels 204; Gaps 25;
Matches 110; Conservative 72; Mismatches

QY 11 AYLVGSSSHVMAKLGARQRTTLPPL-----TNFPGQAPLPTLTOHTTE 54
DQ 1245 AVVVLQTSKAEIIGNGVDETTLTATVDPDNAMVKDQVTFSTN-PADQLQSQSKSTND 1303
QY 55 SGAND-----TILPLANNIQGLVGMKKQKREPVFQVNDATSEFTALKTYVQRTS-- 107
DQ 1304 SGVAEVTFFKGVLTGVTAEATLTPWGNNDTK---IVNIAPASNAQVTLNIPAOQVVTNNS 1360
QY 108 ---AALISDPGQGPLAFVNLGFS-----NTGL---QALG----- 136
DQ 1361 DSVQLTATVADPSNHPAGITVNTFMPDVAAANFTLENNGIALTOANGEAHVTLKKGKAG 1420
QY 137 ---ITDDLGAQFPDGPADAAANIGDLSQVVA-----PFTGTTINGVFLIGSDODFL 187
DQ 1421 THVTATLSNNNTSDQ--PVTFAADKTSALVVLQISKNETLNGVDSATLTATVXQDF 1477
QY 188 DQFTDDISSTFGSSITVOVALSGSARPPDQAGNEHFGFLDGISQPSVTGMETTVPGQAV 247
DQ 1478 DNEVNNLPVTFSTASSGLTLTPGESNTNE-----SGIAQATLAG----- 1516
QY 248 VPRGIIILGRDGDGTGRPSNALDSFMAFRHFOQKVEFNAAYTLNANIPANSAGNLTOE 307
DQ 1517 -----VAFEGVTVASLANNGA---SDNK 1537
QY 308 GAEPFGARMGRMGSGAPILDAPADDPALGADPQRNNNDYSDTLTDETRCP----- 361
DQ 1538 TVHITIGT-----AAAKIETLTPVPSDIAGT-PQNSSGSVITATVDNNGFPKGVTVN 1591
QY 362 ---GAHVKTNPPO---DLGSPVDTFHAMRSSIPYG--PETSDAELASG----- 402
DQ 1592 FTSNAATAEMTNGQAVTNEQKATVTVNTRSSIESGARPDVTEASLENGSSTLSTSN 1651
QY 403 -----VTAODRGLLFVEYOSIIGNGFRFOQIMANNANPFPSKPTT 443
DQ 1652 VNADASTAHLTLQLPDTVSAGERTSLYIEVKDNYNGVPGQEVTLASVP-----SEGV 1707
QY 444 PGIPPIIQOTTPRTVGGDLPLNQNFTFVPLFVIRKGBEY-----PFLPSISA 496
DQ 1708 PSNNAI-YTT-----NHDGNFYAS--FTATKAGVY-----QVTAITL 1740

RESULT 4
LKA3_PASHA
ID_LKA3_PASHA STANDARD; PRT; 953 AA.
AC P55116;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Leukotoxin from serotype T3.
GN LKA.
OS Pasteurella haemolytica.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OX Mannheimia.
OX NCBI_TaxID=75985;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=serotype T3;
RA MEDLINE=94041617; PubMed=8225575;
RX Burrows L.L., Lo R.Y.C., Olah-Winfield E.;
RT "Molecular analysis of the leukotoxin determinants from Pasteurella
RT haemolytica serotypes 1 to 16."
RL Infect. Immun. 61:5001-5007(1993).
CC -!- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD
CC CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY
CC DEFINED.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: THE GLY-RICH REGION IS PROBABLY INVOLVED IN BINDING
CC CALCIUM, WHICH IS REQUIRED FOR TARGET CELL-BINDING OR CYTOLYTIC
CC ACTIVITY.
CC -!- DOMAIN: THE THREE TRANSMEMBRANE DOMAINS ARE BELIEVED TO BE
CC INVOLVED IN PORE FORMATION BY THE CYTOTOXIN (BY SIMILARITY).
CC -!- PTM: PALMITOYLATED BY LKTC. THE TOXIN ONLY BECOMES ACTIVE WHEN
CC MODIFIED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE RTX PROKARYOTIC TOXIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U01216; AAB36691.1;
DR InterPro: IPR001343; HemIysn_Ca bind.
DR InterPro: IPR003355; RTXtoxin_N.
DR InterPro: IPR003995; RtxA.
DR Pfam: PF00353; hemolysinCbind; 5.
DR Pfam: PF02382; RTX; 1.
DR PRINTS: PR003113; CABNDNGRPT.
DR PRINTS: PR01488; RTXTOXINA.
DR PROSITE: PS00330; HEMOLYSIN_CALCIUM; 2.
KW Hemolysis; Toxin; Cytolysis; Cytotoxin; Repeat; Calcium;
KW Transmembrane; Lipoprotein; Palmitate.
FT TRANSMEM 229 249
FT TRANSMEM 297 318
FT TRANSMEM 381 401
FT DOMAIN 734 784
FT REPEAT 734 739
FT REPEAT 752 757
FT REPEAT 761 766
FT REPEAT 770 775
FT REPEAT 779 784
FT REPEAT 779 784
SQ SEQUENCE 953 AA; 101948 MW; FDBDCB2FDC85FDF2 CRC64;

Query Match
Best Local Similarity 4.2%; Score 108.5; DB 1; Length 953;
Matches 99; Conservative 60; Mismatches 146; Indels 169; Gaps 25;

QY 26 GAROTRT-TPLLNFPQAPLPLTQ--HTTESGANDTILPL-----NNIQGDI 71
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 11 GIRSTLTATRGGLNRAQQ---SLTQAGOTLKNGVKKILYIPKDYKDSGNGLO-DL 65

QY 72 L-----VGMKKOKERFVFQVNDATSPKTKATYVVPRI---TSAATLISDPQQPLAFV 123
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 66 VKAAEELGIEVQKE-----EGNDIAKAQTSGLTI--QNVIGLTERGIVLSAPOLDKL--- 115

QY 124 NLGFSNTGLQALGTTDDLGAQFPDQGFADAANLGGDLSQWAVFTGT-TIHGVFLIGSD 182
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 116 -LQNKVG-QALGSSE-----STQNFQSAKTVLSGVQVILSGVLAGMD 157

QY 183 QDD-----FLDQFTDDISSTFGSSITOVQALSG- 210
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 158 LDEALQNESDQLTAKAGLELTNSLIENIANSVQTLDAFSEIQS-FGSKLQNVKGLGAL 216

QY 211 -----SARPPDAQG---HEHFGLDGLSQPSVGTGWETTVPFGQAVVPPGIIILTRGDGTG 262
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

DB 217 GDKLKNIGGLDKAGLGLHVISGLSGAT-----AALVLADKDATA 257
QY 263 TRPSWALDGSFMAFRHFOQKVPFNAFTYLANAIPANSAGNLTOQEGAEFLGARMFORWKS 322
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 258 KKVQ-----AGFELANQV-----VGNITKAVSSYILAQRVAARLSS 293
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 323 GAPIDLAPTADDPALGADPQR-----NNNEDYSDTLTDETRCFPGAHVRTNPRQDLGGPV 378
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 294 TGPV-AALIASTAVTALAIPLSPFAGIADKPDRAKLUEN-----YAEFRKGLGVEGD----- 342
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 379 DTFHAMRSIPYGPETSDAE-----LASGVTAQDRGLLFFVEYQSIINGG 422
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 343 ----SLLAEYQGTGTIDASVTAINALAAAGVSAAGSVVSPALLVSG 392
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 5
NANH MICVI
ID NANH MICVI STANDARD; PRT; 647 AA.
AC Q02834;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sialidase precursor (EC 3.2.1.18) (Neuraminidase).
GN NEDA.
OS Micromonospora viridifaciens.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Micromonosporineae; Micromonosporaceae;
OC Micromonospora.
OX NCBI_TaxID=1881;
RN (1)_TaxID=1881;
RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-647.
RC STRAIN=ATCC 31146;
RX MEDLINE=93015752; PubMed=1400240;
RA Sakurada K., Ohta T., Hasegawa M.;
RT "Cloning, expression, and characterization of the Micromonospora
RT viridifaciens neuraminidase gene in Streptomyces lividans.";
RL J. Bacteriol. 174:6896-6903 (1992).
RN (2)
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RC STRAIN=ATCC 31146;
RX MEDLINE=9616436; PubMed=8591030;
RA Gaskell A., Crenell S., Taylor G.;
RT "The three domains of a bacterial sialidase: a beta-propeller, an
RT immunoglobulin module and a galactose-binding jelly-roll.";
RL Structure 3:1197-1205 (1995).
CC -!- FUNCTION: TO RELEASE SIALIC ACIDS FOR USE AS CARBON AND
CC ENERGY SOURCES FOR THIS NON-PATHOGENIC BACTERIUM WHILE IN
CC PATHOGENIC MICROORGANISMS, SIALIDASES HAVE BEEN SUGGESTED TO
CC BE PATHOGENIC FACTORS.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2->3)-, alpha-(2->6)-,
CC alpha-(2->8)-glycosidic linkages of terminal sialic residues in
CC oligosaccharides, glycoproteins, glycolipids, colominic acid and
CC synthetic substrates.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- INDUCTION: BY N-ACETYLNEURAMINIC ACID, COLOMINIC ACID, AND SIALIC
CC ACID.
CC -!- SIMILARITY: BELONGS TO FAMILY 33 OF GLYCOSYL HYDROLASES.
CC -!- SIMILARITY: CONTAINS 7 BNR REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D01045; BAA00852.1;
DR PIR: A45244; A45244.
DR PDB: 1EUR; 11-JAN-97.
DR PDB: 1EUS; 11-JAN-97.
DR PDB: 1EUT; 11-JAN-97.
DR PDB: 1EUU; 11-JAN-97.

```



```

DR InterPro: IPR000421; FA58.C.
DR InterPro: IPR002860; GH_BNR.
DR Pfam: PF02012; BNR. 5.
DR SMART: SM00231; FA58C. 1.
KM Hydrolase; Glycosidase; Signal; Repeat; 3d-structure.
FT SIGNAL 1 37
FT CHAIN 38 647
FT ACT_SITE 260 260 NUCLEOPHILE.
FT ACT_SITE 370 370 NUCLEOPHILE.
FT REPEAT 102 113 BNR 1.
FT REPEAT 175 186 BNR 2.
FT REPEAT 239 250 BNR 3.
FT REPEAT 287 298 BNR 4.
FT REPEAT 348 359 BNR 5.
SQ SEQUENCE 647 AA; 68830 MW; DCCIFESBC935B8AD CRC64;

Query Match 4.28; Score 108; DB 1; Length 647;
Best Local Similarity 20.18; Pred. No. 1.77; Indels 220; Gaps 29;
Matches 110; Conservative 48; Mismatches 169;

QY 79 KERFVFQVNDATSPKALTKYVPPQRTISAAILSDPSQPL-----AFVN--LGFSTNG 131
DB 12 RRRVAVFLPALALAAIYVAGSPQAALNAGAV---PEGGEFLYTEQDLAVNGREGCPMYR 68
QY 132 LQALGITDDIGDAQPPDGFADANLGDLSQWVAPFTTTHGVFLIGSDQDDFLDQFT 191
DB 69 IPALTVPD-----GDLASVDGRPTGIDAPG-----PNSILGRS 104
QY 192 DDISSTGSSITVOALSGSARPPDQAGHEFGLDGISQS-VTGMET-TVF----- 242
DB 105 TDGRTWGEQ--QVVASGQTTAP-----IKGFSDSYLVDRETGTIFNFHYVSO 151
QY 243 -----PGQAVVPPGI---ILTGRD-----DTGRPSMALDGSFMA 275
DB 152 RQGRAGSRPGTDPADPNVLHANVATSDGILTMSHRTITADITDPGRKSRFAASGEIQ 211
QY 276 FRHFQKVPFNAYLTLANIPANSAGNLTQEGAEFLGARMGR-WMSGADIDLAPTAD 334
DB 212 LRYPHAGRLIQYTIINAGAPQAVSVYSD-----HGRTWRAG----- 251
QY 335 PALGADQRNNNFYSD-----TLTDERCPFGAHRKTNRODLGCVDFTFHMRSSIP 389
DB 252 EAVGVGNDENTKVELSGRVLNRSRDSRSGY---RVAVSTD-GG-----HS----- 295
QY 390 YGPESTDAEL-----ASGVTADRGLLF----- 412
DB 296 YGPVITDRDLDPDNNNAIIPAPDAPARAKVLLPSNAAOTSRSQGTIRMSCDDCQ 355
QY 413 -----VEYOSII-----GNGFRFOQIN--WANNANPFPSKP---I 442
DB 356 TWPVSKVPQPSMSYSTITALPDGTGYLLYEPGTGIRANNNLWLGICAPFTIPDVAL 415
QY 443 TPGIEPII-----GQTRP-----TVGLDPLNONE-----TFTVPLEV 476
DB 416 EPGQGVTVPAVTVNQSGIAPVKPSLIQDASPDWQGVGSVEPLMGROAKGQVITIVPAGT 475
QY 477 IPRKGEY 483
DB 476 TP--GRY 480

RESULT 6
PGCV CHICK
ID PGCV CHICK STANDARD; PRT; 3562 AA.
AC Q90953; Q90945;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Versican core protein precursor (large fibroblast proteoglycan)
DE (chondroitin sulfate proteoglycan core protein 2) (Pg-M).
GN CSRG2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Archaeosuria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=white leghorn; TISSUE=Limb bud;
RX MEDLINE=93300846; PubMed=8314802;
RA Shinomura T., Nishida Y., Ito K., Kimata K.;
RT "CDNA cloning of PG-M, a large chondroitin sulfate proteoglycan
RT expressed during chondrogenesis in chick limb buds. Alternative
RT spliced multiforms of PG-M and their relationships to versican.";
RL J. Biol. Chem. 268:14461-14469(1993).
CC - FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC - SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC - ALTERNATIVE PRODUCTS: At least 2 isoforms; V0 (shown here) and
CC V1; are produced by alternative splicing.
CC - TISSUE SPECIFICITY: Prechondrogenic condensation area of
CC developing limb buds.
CC - DEVELOPMENTAL STAGE: Disappears after the cartilage development
CC (by similarity).
CC - SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC - SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC - SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC - SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC - SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X60226; CAA42787.1; -
DR EMBL; D13542; BAA02742.1; -
DR HSSP; P00740; 1EDM.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF-2.
DR InterPro: IPR001881; EGF_CA.
DR InterPro: IPR003598; Ig_2.
DR InterPro: IPR003066; Ig_MHC.
DR InterPro: IPR001304; Lectin_C.
DR InterPro: IPR000538; Link.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF_2.
DR Pfam; PR00047; Ig_1.
DR Pfam; PR00059; lectin_C_1.
DR Pfam; PF00084; sushi_1.
DR Pfam; PF00193; Xlink_2.
DR Prodom; PD000918; Link_2.
DR SMART; SM00032; CCP_1.
DR SMART; SM00034; CLECT_1.
DR SMART; SM00179; EGF_CA_1.
DR SMART; SM00001; EGF-like_1.
DR SMART; SM00409; Ig_1.
DR SMART; SM00445; Link_2.
DR PROSITE; PS00010; ASX_HYDROXYL_1.
DR PROSITE; PS00022; EGF_1_2.
DR PROSITE; PS01186; EGF_2_1.
DR PROSITE; PS01187; EGF_CA_1.
DR PROSITE; PS01241; Link_2.
DR PROSITE; PS00615; C-TYPE LECTIN_1_1.
DR PROSITE; PS00645; C-TYPE LECTIN_2_1.
KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
KW Hyaluronic acid; Alternative splicing.
FT SIGNAL 1 26
FT CHAIN 27 3562
FT VERSICAN CORE PROTEIN.

```

```

FT DOMAIN 37 136 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 166 243 LINK 1.
FT DOMAIN 264 345 LINK 2.
FT DOMAIN 3254 3290 EGF-LIKE 1.
FT DOMAIN 3292 3328 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3341 3455 C-TYPE LECTIN.
FT DOMAIN 3460 3518 SUSHI.
FT DISULFID 44 129 BY SIMILARITY.
FT DISULFID 171 242 BY SIMILARITY.
FT DISULFID 195 216 BY SIMILARITY.
FT DISULFID 269 344 BY SIMILARITY.
FT DISULFID 293 314 BY SIMILARITY.
FT DISULFID 3258 3269 BY SIMILARITY.
FT DISULFID 3263 3278 BY SIMILARITY.
FT DISULFID 3280 3289 BY SIMILARITY.
FT DISULFID 3296 3307 BY SIMILARITY.
FT DISULFID 3301 3316 BY SIMILARITY.
FT DISULFID 3318 3327 BY SIMILARITY.
FT DISULFID 3334 3345 BY SIMILARITY.
FT DISULFID 3362 3454 BY SIMILARITY.
FT DISULFID 3430 3446 BY SIMILARITY.
FT DISULFID 3461 3504 BY SIMILARITY.
FT DISULFID 3490 3517 BY SIMILARITY.
FT CARBOHYD 163 163 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 235 235 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 329 329 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 529 529 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 709 709 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 948 948 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1409 1409 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1479 1479 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1523 1523 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1530 1530 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1625 1625 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1751 1751 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1988 1988 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2088 2088 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2089 2089 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2507 2507 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2642 2642 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2679 2679 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2748 2748 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2762 2762 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3069 3069 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3194 3194 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3232 3232 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3545 3545 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPIC 485 1411 MISSING (IN ISOFORM VI).
SQ SEQUENCE 3562 AA; 388078 MW; 98C56E88C1602D2 CRC64;

Query Match 4.1%; Score 107; DB 1; Length 3562;
Best Local Similarity 22.1%; Pred. No. 20;
Matches 92; Conservative 46; Mismatches 159; Indels 120; Gaps 22;

QY 47 TL70----HTTEGANDTILPLANNIOGDIIVGMKKOKERVFVQVNDATSFKALTXYVP 102
Db 984 TMTEGGQISVTSAEKESVAALQEREPQSVGLPETKEPKFTDV-----TEIETTPV 1036
QY 103 QRTISAAILSDPSQOPLAFVNLGFSNTGLQALGITDLDGADFPDQGPADAANLGDLS 162
Db 1037 OR-----EGDTSVLP-----VTVG-----SEDIGEMQVTHTSFDS-----II 1069
QY 163 QWVAPFTGTHGVF---LIGSQDDFLDQFTDDISSTFGSS---ITQVQALSGSARPFQD 217
Db 1070 HTEATVTSKASEVFPKELSTKQD-----RELGTAMGSTLPVTSVQV----- 1112
QY 218 AGHEHFGFLDGIQSPSVTGW-----TTVFPGQAVVPPGIIILT-----GRDGTGT 263
Db 1113 --HEQ-KTTAGFESPOTTQKHDEMGMSAYDEMPATELSVPALMLTEYQVSGPVETST 1169
QY 264 RPSWALDGSFMAFRHFOOKVPEFNAYVLANAIPA-----NSAGNLTOQSGAEPLGA 314
Db 1170 R-SLHLTGTPKAETATDQE-----EKITEAVPVTFGTQAKVYESKGTTRRED----- 1216
```

```

QY 315 RMFGRWKSAPIDLAPTADDPA-----LGADP-QRNNNFDYSDTLTDETCCPFGAHVR 366
Db 1217 RDVGSMNSVLPHTMLSPSTAGSISLLTLGASPSQTPEGSISBELEEVKTVPFSSRAT 1276
QY 367 -KTNPRQDLG---GVDTFHAMRSGISPYG-----PETSDAELASGVTAQDRL 410
Db 1277 DKTIVISDITSSISNVDKIQPTSASKPFVSSKSPRIIPEDEEVTSSDIIVIDESI 1333

RESULT 7
N121 HUMAN STANDARD; PRT: 1229 AA.
ID Q9Y2N3; O75115; Q9Y4S7;
AC 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nuclear envelope pore membrane protein POM 121 (Pore membrane protein
of 121 kDa) (P145).
GN NUP121 OR KIAA0618.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
OX [1] SEQUENCE FROM N.A.
RN Cordes M., Bauer C., Holmes A.;
RA Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 243-1229 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
RN [3]
RP SEQUENCE OF 1130-1229 FROM N.A.
RC TISSUE=Uterus;
RA Koehler K., Beyer A., Mewes H.-W., Gassenhuber J., Wiemann S.;
RA Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ESSENTIAL COMPONENT OF THE NUCLEAR PORE COMPLEX. THE
REPEAT-CONTAINING DOMAIN MAY BE INVOLVED IN ANCHORING COMPONENTS
OF THE PORE COMPLEX TO THE PORE MEMBRANE (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. NUCLEAR PORE
MEMBRANE (BY SIMILARITY).
CC -!- DOMAIN: CONTAINS F-X-F-G REPEATS.
CC -!- SIMILARITY: THE REPEAT REGION COMPOSED OF PENTAPEPTIDE REPEATS
SEPARATED BY SER/THR-RICH DOMAINS IS SIMILAR TO THAT OF YEAST
NUP1, NSP1 AND MAMMALIAN P62 AND NUP153.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AC006014; AAC28064.1; -
DR EMBL; AB014518; BAA31593.1; -
DR EMBL; AL080109; CAB45713.1; -
KW Nuclear protein; Transport; Transmembrane; Repeat.
FT DOMAIN 1 40 CISTERNAL SIDE (POTENTIAL).
FT TRANSMEM 41 61 POTENTIAL. PORE SIDE (POTENTIAL).
FT DOMAIN 62 1229 PORE SIDE (POTENTIAL).
FT DOMAIN 4 10 POLY-ALA.
FT DOMAIN 51 56 POLY-ALA.
FT DOMAIN 294 299 POLY-LYS.
FT DOMAIN 441 444 POLY-SER.
FT DOMAIN 499 502 POLY-PRO.
```

```

FT DOMAIN 733 736 POLY-SER.
FT DOMAIN 819 826 POLY-SER.
FT DOMAIN 869 875 POLY-THR.
FT DOMAIN 1061 1067 POLY-SER.
FT CONFLICT 297 297 K -> E (IN REF. 2).
FT CONFLICT 336 336 S -> N (IN REF. 2).
FT CONFLICT 379 379 T -> A (IN REF. 2).
FT CONFLICT 451 451 K -> R (IN REF. 2).
FT CONFLICT 545 545 T -> P (IN REF. 2).
FT CONFLICT 559 559 S -> P (IN REF. 2).
FT CONFLICT 702 702 S -> P (IN REF. 2).
FT CONFLICT 881 881 P -> H (IN REF. 2).
FT CONFLICT 965 965 T -> A (IN REF. 2).
FT CONFLICT 982 984 PAT -> A (IN REF. 2).
FT CONFLICT 991 996 ASTRIKI -> PSMKV (IN REF. 2).
FT CONFLICT 1000 1000 H -> Y (IN REF. 2).
FT CONFLICT 1006 1008 OPT -> HPI (IN REF. 2).
FT CONFLICT 1044 1044 MISSING (IN REF. 2).
FT CONFLICT 1102 1102 A -> T (IN REF. 2).
FT CONFLICT 1124 1124 T -> A (IN REF. 2).
FT CONFLICT 1146 1146 S -> G (IN REF. 1).
FT CONFLICT 1165 1165 L -> Q (IN REF. 1).
FT CONFLICT 1195 1195 A -> G (IN REF. 1).
FT CONFLICT 1199 1229 SAAPSFISGSKTPGARQRLQAROHTRKK -> NTEAHQ
OEHSPRKPNLSRKLLPAVRAQGPFRGQASSFPRKE
(IN REF. 2).
P -> L (IN REF. 3).
CONFLICT 1202 1202
FT SEQUENCE 1229 AA; 125087 MM; 515655D1285898B CRC64;

Query Match 4.0%; Score 105; DB 1; Length 1229;
Best Local Similarity 22.9%; Pred. No. 6.7;
Matches 83; Conservative 35; Mismatches 153; Indels 92; Gaps 17;

QY 91 TSFPTALKTVPPQRTGAA-----ILISDP--SQGLAFVNLGFSNGLQALGTD 139
DB 874 TTFSQSLPTAVPRTATSSAADPSGFGSTLATSPATSSQP---TLTFSNTSTPTFNI-- 927
QY 140 DLGDAQPPDQGFADANLGDLLSQWVAFPTGTTIHGVFLI-----GSDQDDFLDQFTDD 193
DB 928 -----PFSSAKSP-----LPSYPCANQPQAFGAEGCPQPAKAKALPSPGCS- 970
QY 194 ISSFTGSSITVOALSSGARP-----PDQAGHEHGF-----LD 227
DB 971 -SFTFGNSAAPAPAPATPAPASTIKIVPAHVPTPIQTFPGATHSAFGLKATASAFGAP 1029
QY 228 GISQPSVTGWTTFPPQAVVPPGIIILGRDGDGTSPSMLDSSFMARFRHQKVEPFN 287
DB 1030 ASSQPAF-CGSTAVFSFGAATSGFGATTTQASSSSSS-----VFGSTTP--S 1075
QY 288 AYT-L-ANNAIPANSAG--NLTOQGAEFILGARMFGRMKSGAPIDLAPTA---DDPALGAD 340
DB 1076 PFTTGGSAAPAGSSFGINVTTPSSSATTTGAFSFGAGOGSGTATSTPTGGLGNALGTT 1135
QY 341 PQRNNNFDDYDSTLTDETRCPGAHVKTNPRLDGGPVDFTHAMRSSIPLYGPETSDAELA 400
DB 1136 GQ-STPFAFNVNSTTESKPVFG---TATPFTGLNTPAPGVGTSSSLSPASSAPAGGF 1191
QY 401 SGV 403
DB 1192 VGV 1194

RESULT 8
NU42_YEAST STANDARD; PRT; 430 AA.
AC P49686;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nucleoporin NUP42 (Nuclear pore protein NUP42).
GN NUP42 OR R1P1 OR YDR192C OR YD9346.04C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

```

```

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=95360993; PubMed=7634338;
RA Stutz F., Neville M., Rosbash M.;
RT "Identification of a novel nuclear pore-associated protein as a
functional target of the HIV-1 Rev protein in yeast.";
RL Cell 82:495-506 (1995).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=528bc / AB972;
RA Oliver K., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INTERACTS SPECIFICALLY WITH THE HIV-1 REV PROTEIN
EFFECTOR DOMAIN AND PROMOTES RNA EXPORT.
CC -!- SUBCELLULAR LOCATION: Nuclear pore complex.
CC -!- DOMAIN: CONTAINS F-G REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U30614; AAA8703.1; -.
DR EMBL; Z48784; CAAB8706.1; -.
DR SGD; S0002600; NUP42.
DR InterPro; IPR004325; Nucleoporin FG.
DR Pfam; PF03093; Nucleoporin FG; 26.
KW Nuclear protein; Transport; Repeat.
FT CONFLICT 331 331 K -> Q (IN REF. 1).
FT CONFLICT 419 419 V -> D (IN REF. 1).
SQ SEQUENCE 430 AA; 42778 MM; 357E94914A5261F4 CRC64;

Query Match 4.0%; Score 104.5; DB 1; Length 430;
Best Local Similarity 22.4%; Pred. No. 1.8;
Matches 94; Conservative 38; Mismatches 154; Indels 134; Gaps 20;

QY 138 TDDLGDAQPPDQGFADANLGDLLSQWVAFPTGTTIHGVFLISDQDDFLDQFTDDISDT 197
DB 32 TNNMGSAFGRPSGCTANTM-----TGCTTSAF-----GMPQFGTNTGNT 72
QY 198 FGSSITVOALSSGARPFDQAGHEHGFLLD--GISQPSVTGWTTFPPQAVVPPGIIILT 255
DB 73 GNTSISAFGNTSNAAKP-SAFGAPAFGSSAPINVPSTT-----SAFGAPSP-- 119
QY 256 GRDDDTGTRPSMLDGSFMAFRHQKQVPEFNAVTLNATIPANSAGNLTOQGAEFILGAR 315
DB 120 ---GSTG-----FGAV-----AATSNPFSGSPSGMSAF--GQP 148
QY 316 MFGRMKSGAPIDLAPTADDPALGA-----DPQRNNFVDSYSDT----- 352
DB 149 AFGANKTAIRISSSVSNNSNNAFGAASNTPLTTTSPFGLQNMSONNSSATSGKPTFG 208
QY 353 LTDETRCPGAHVKTNPRLDGL-GPVDTF-----HAMRSSIPLYGPETSD 396
DB 209 AATNTQSPFGT-IQNTSTSGTGVSFPGFTGTSNNKSPFSNLQSGAGASSPFGTTTSK 267
QY 397 AELASGVTAQDRGLLFVEYQSIIGNGFRF---QOINANNANFP-----FSKPIITGI 446
DB 268 ANNNNNVGSAPFTTNQSPSGSGGTFSASNLNKNATGNFQSSGKNGFSGLTIPON 327
QY 447 E-----PILGQTPRTVGLDP---LNQNETFTVPLFVIPKGEYFFLPISALTAT 495
DB 328 DANKVSGSNSFGGTMTNT---DPNLSLKSNGNATISFGF---GQOQNMATNNATATAT 379

RESULT 9
BPRV_BACNO STANDARD; PRT; 603 AA.
ID BPRV_BACNO

```



```

RA  Servin M.S., Mate M.J., Switala J., Fita I., Loewen P.C.;
RT  "Role of the lateral channel in catalase HPII of Escherichia coli.";
RL  Protein Sci. 8:490-498 (1999).
RN  [6]
RP  X-RAY CRYSTALLOGRAPHY (1.89 ANGSTROMS).
RX  MEDLINE=21348730; PubMed=11455600;
RA  Melik-Adamyan W.R., Bravo J., Carpena X., Switala J.,
RA  Mate M.J., Fita I., Loewen P.C.;
RT  "Substrate flow in catalases deduced from the crystal structures of
RT  active site variants of HPII from Escherichia coli.";
RL  Protein 44:270-281 (2001)
CC  -1- FUNCTION: DECOMPOSES HYDROGEN PEROXIDE IN WATER AND OXYGEN, SERVES
CC  TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN PEROXIDE.
CC  -1- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
CC  -1- COFACTOR: HEME GROUP.
CC  -1- SUBUNIT: HOMOTETRAMER.
CC  -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC  -1- INDUCTION: BY ENTRY INTO STATIONARY PHASE.
CC  -1- SIMILARITY: BELONGS TO THE CATALASE FAMILY. HPII SUBFAMILY.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; M55161; AAA24039.1; -.
DR  EMBL; AE000268; AAC74802.1; -.
DR  EMBL; D90815; BAA20916.1; -.
DR  EMBL; D90816; BAA15513.1; -.
DR  EMBL; D90817; BAA15521.1; -.
DR  PIR; A39129; A39129
DR  PDB; 11PH; 04-SEP-97.
DR  PDB; 1CE9; 06-APR-99.
DR  PDB; 1QF7; 26-APR-99.
DR  PDB; 1GQ9; 30-AUG-00.
DR  PDB; 1GGE; 30-AUG-00.
DR  PDB; 1GGF; 30-AUG-00.
DR  PDB; 1GGH; 30-AUG-00.
DR  PDB; 1GGJ; 30-AUG-00.
DR  PDB; 1GGR; 30-AUG-00.
DR  EcoGene; EG10509; katE.
DR  InterPro; IPR002226; Catalase.
DR  Pfam; PF00199; catalase; 1.
DR  PRINTS; PR00067; CATALASE.
DR  ProDom; PD000510; Catalase; 1.
DR  PROSITE; PS00437; CATALASE_1; 1.
DR  PROSITE; PS00438; CATALASE_2; 1.
KW  Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide;
KW  3D-structure; Complete proteome.
FT  ACT_SITE 128 128 BY SIMILARITY.
FT  ACT_SITE 201 201 BY SIMILARITY.
FT  METAL 415 415 IRON (HEME PROXIMAL LIGAND).
FT  CONFLICT 198 198 L -> F (IN REF. 2).
SQ  SEQUENCE 753 AA; 84162 MW; 1F034E486A70FB9 CRC64;
Query Match 4.0%; Score 104; DB 1; Length 753;
Best Local Similarity 19.9%; Pred. No. 4.1;
Matches 123; Conservative 55; Mismatches 205; Indels 234; Gaps 31;
QY 40 PGQAPLPLTHTTBSG--ANDTILPLNIGDILVKKKKKKERFVFQ-----VNDAT 91
DB 38 PAAPETPGAPAPAPGSLKAPDT--RNEKLNLSLEDVRKSEVYALTTNGVGRIADQON 93
QY 92 SFK-----TALKTY-----VPORTSA-----ATLI 112
DB 94 SLRAGSRRPTLLDFILREKXITTHDHERIPRIYAABSAAHGYFQPKSLDITKDFL 153
QY 113 SDPSQOQPLAFVNL-----GFSNTGLOALGITDDLG-----DA-Q 145
DB 154 SDPKITPVEFVRFSVGGAGSADTVRDIRGFATKFTYEEGIFDLVGNNTPIFFIQDAMK 213

```

```

QY 146 PDGQFADANLGDLDLSQWVAPRTGTHGVLIGSDODDPLDQFTDIDISSTFGSSITOV 205
DB 214 FPD--FVHAVK---PEPHWAI-P-QGOSAH-----DTFMDYVSLQETELHNW 254
QY 206 QALSGSARPPDQAGHEHFGF-----LDGI SOPSTVGTWETTPPGQAVVPPGIIITGRD 258
DB 255 WAMSDRGIPRNSYRTMGFFGHTFRLINAEKATFVRFHMKPLAGKASLVWDEAQKLTGRD 314
QY 259 GDTGRPSW-ALD-GSPFMARHFQKVPEFNATL-----ANAI PANSAAGNLT 304
DB 315 PDHRLRLMELAIADGDPPEYELGPULPESEDFKFDLIDPTKLIPEELVPOVRGKMV 374
QY 305 QQSGAEFLGARMGWRMKSQA--PIDLAPTD---DPALGADPORNNNFDSDTLTDETRC 359
DB 375 LNNRPD---NFAENBQAAFHGHIVPGIDFTNDPL-----QGRIFSVDIOISR--- 422
QY 360 PGGAHVAKTNPRDGLGPDVTFHAM--RSSIFY-----GPEISDA----- 397
DB 423 -----LGGP--NFHEIPIINRPTCPYHNFORDGMRMGIDTNPANVEPNSIN 466
QY 398 -----ELASGVTADRGLLFVEYOS-----IIGNGRFPQOINMAN 433
DB 467 DNMRETPPGPKRGFESYGERVEGNKVRERSFSFGELYHPLFLWLSQTFEQRHIVDG 536
QY 434 ANFPFSKPTIPGI-EPIIQG-----TTPTVIGL--DP 463
DB 527 FSPFLSKVNPYPIRERVVDLAHIDLTLAGAVAKNIGIELTDQNLITPPDVNGLKKDP 566
QY 464 LNONETFTVEPLFVIPKG 480
DB 587 -----SLSLVAIPDG 596
RESULT 11
ID FOXA_YEREN STANDARD; PRT; 710 AA.
AC 001674;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Ferrioxamine receptor precursor.
GN FOXA.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51872 / WA-C / Serotype O:8;
RX MEDLINE=92349959; PubMed=1640832;
RA Baeninger A.J., Hanke K.;
RT "Ferrioxamine uptake in Yersinia enterocolitica: characterization of
RT the receptor protein Foxa."
RL Mol. Microbiol. 6:1309-1321 (1992).
RN [2]
RP REVISIONS.
RC STRAIN=ATCC 51872 / WA-C / Serotype O:8;
RA Baeninger A.J.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: FERROXAMINE BINDING AND UPTAKE, IN ASSOCIATION WITH THE
CC TONB PROTEIN.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

DR EMBL: X60447; CAA42975.1; --
DR PIR: S22673; S22673.
DR HSSP: P06971; IQJQ.
DR InterPro: IPR00531; TonB_boxC.
DR Pfam: PF00593; TonB_boxC; 1.
DR PROSITE: PS00430; TonB_DEPENDENT_REC_1; 1.
DR PROSITE: PS01156; TonB_DEPENDENT_REC_2; 1.
KW Signal; Receptor; Transmembrane; Outer membrane; Iron transport;
transport; TonB_box.
FT SIGNAL 1 26
FT CHAIN 27 710
FT SITE 28 35
FT SITE 693 710
FT SITE TONB C-TERMINAL BOX.
FT TRANSMEM 29 37
FT TRANSMEM 65 73
FT TRANSMEM 91 99
FT TRANSMEM 106 114
FT TRANSMEM 137 145
FT TRANSMEM 152 160
FT TRANSMEM 180 188
FT TRANSMEM 194 202
FT TRANSMEM 208 216
FT TRANSMEM 259 267
FT TRANSMEM 271 279
FT TRANSMEM 293 301
FT TRANSMEM 309 317
FT TRANSMEM 353 361
FT TRANSMEM 370 378
FT TRANSMEM 427 435
FT TRANSMEM 443 451
FT TRANSMEM 476 484
FT TRANSMEM 491 499
FT TRANSMEM 517 525
FT TRANSMEM 531 539
FT TRANSMEM 555 563
FT TRANSMEM 567 575
FT TRANSMEM 579 587
FT TRANSMEM 610 618
FT TRANSMEM 624 632
FT TRANSMEM 649 657
FT TRANSMEM 671 679
FT TRANSMEM 684 692
FT TRANSMEM 702 710
FT SEQUENCE 710 AA; 78382 MW; 82EB6EC1546900B8 CRC64;
Query Match 4.0%; Score 103.5; DB 1; Length 710;
Best Local Similarity 21.0%; Pred. No. 4.1;
Matches 118; Conservative 61; Mismatches 215; Indels 167; Gaps 23;
QY 62 LPLNNIGD-ILVGMKKQKRFVFFQVNDATSPKALTATVYQRIITSAALISDPQ--- 117
DB 20 IPLASIAADTTIEVTAKAGHEADLPTSGYTATTTKGATKTDQPLILTAQSVSVVTRQOMD 79
QY 118 QPLAFVNLGFSNTGLQALGITDLDGDAQFPDGFADANLGGDLSQWVAPFTGTHGVF 177
DB 80 QNVATVN-----QALNTPGVFTG-----FSGATRYDIVA--LRFGHGDVNTF 123
QY 178 LIG-----SDQDDF-----LDQFTDDISSTFGSSITQVQALSGSARP-FDQAGH 220
DB 124 LDGLRLLSDGGSYNVLQVDPWFLERIDVIKGPSSALYGQSI PGVVMWMTSKRPQFTSEGH 183
QY 221 -----EHGFG-LDGISQPSVTGHTVTPQQAUVV----- 249
DB 184 FRLTAGNNNTQVAAFDYDTDAISEHWFRLTGITRNSDTMYDHOREEYAIAPSLLMQDPE 243
QY 250 -PGIILTG---RGDCTTRPSMALDGSFM-----APRFQOKVPERNAV 289
DB 244 NTSLLLRANLQKQPSGYHSAVPADGSIYQKLSRGFFDGESNNHNVFRWQQ-----IYSY 299
QY 290 TLA-----NAIPANSAGNLTO-QEGAEPFLGARMEFRWKSQ----- 323
DB 300 EFSHKFPDVMSFRQNASYTHSNTQLEQVYQGGWNSDRLNRYVSGEDSSLNFAVDNQL 359

QY 324 -APIDLAPTADDPALGADPQRNNNFYSDTLTDETRCPFGAHRKTNPRQDLGGPVDTH 382
DB 360 EADLRTAAVKHVKVLGVDFQKFRNNLRSDS-----AYATPLNPTVGVGGSTLVS 409
QY 383 AMRSSIYPGTPETSDAEL---ASGVTAQDRGLLFFVEYQSIIGNFRFQQINWANNAN---- 435
DB 410 DYLLTTP-GINTSYLSRRRYEQSGVYLQDQWTLDNWHLNLSGRYDRMKTNINNTANSTDB 468
QY 436 -----FFPSKBITPGIEBIIQOTTPRTVGGLDPLNQNETFTV 472
DB 469 RTDNHASGRASLLYSFDSGISPVSYSAITPSLPDQAQKL-----LKPMT-SSQYEV 521
QY 473 PLFVIPKGEYFPLPSISALT 493
DB 522 GIIYQPPGSTSLYSALYDLT 542
RESULT 12
NUPI_YEAST STANDARD; PRT: 1076 AA.
AC P20676;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nucleoporin NUP1 (Nuclear pore protein NUP1).
GN NUP1 OR YOR098C OR YOR3182C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1] _SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RX MEDLINE=90275616; PubMed=2190694;
RA Davis L.I., Fink G.R.;
RT "The NUP1 gene encodes an essential component of the yeast nuclear pore complex."
RL Cell 61:965-978(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9734368; PubMed=9200815;
RA Voss H., Benes V., Andrade M.A., Valencia A., Rechmann S., Teodoru C.,
RA Schwager C., Paces V., Sander C., Ansorge W.;
RT "DNA sequencing and analysis of 130 kb from yeast chromosome XV."
RL Yeast 13:655-672(1997).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF NUCLEAR PORE COMPLEX.
CC THE PROTEINS MAY BE INVOLVED IN BOTH BINDING AND TRANSLOCATION OF
CC NUCLEOPORINS DURING NUCLEOCYTOPLASMIC TRANSPORT.
CC -!- SUBCELLULAR LOCATION: Nuclear pore complex.
CC -!- DOMAIN: APPEARS TO BE DIVIDED INTO THREE DOMAINS DEFINED BY
CC CENTRALLY LOCATED REPEATING UNITS. FUNCTIONAL N-TERMINAL OF NUP1
CC OR OF NUP2 IS REQUIRED FOR GROWTH.
CC -!- DOMAIN: CONTAINS F-X-F-G REPEATS.
CC -!- SIMILARITY: TO THE CENTRAL REPEATING UNITS OF NSP1 AND NUP2, AND
CC TO MAMMALIAN P62.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M3632; AAA34822.1; -.
CC EMBL: X94335; CAA64020.1; -.
CC EMBL: 275006; CAA99295.1; -.
CC PIR: A35622; A35622.
CC SGD: S0005624; NUP1.
KW Nuclear protein; Transport; Repeat.
FT DOMAIN 333 949 29 X 9 AA APPROXIMATE REPEATS.
SQ SEQUENCE 1076 AA; 113581 MW; 4AC23567D2FB53CC CRC64;
Query Match 4.0%; Score 103; DB 1; Length 1076;


```
Db 802 VAVHASYIEAEVPAE-----TGQTAFL-LKLAGRW-----PVKTIHT----- 842
QY 341 PORNNDYDSDLTDETRCPFGAHRVKNRPNQDLGGPVDTFHAMRSSIPYGPET 394
Db 843 --DNGSNFTST-TVKAACW-----AGIKQEFQ-----IPYNPOS 874

RESULT 14
ID HP12 DEIRA STANDARD; PRT; 1036 AA.
AC P13126;
DT 01-JAN-1990 (Rel. 13, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE Hexagonally packed intermediate-layer surface protein precursor.
GN HPI.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC Deinococaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=Sark;
RX MEDLINE=88032846; PubMed=3667529;
RA Peters J., Peters M., Lottspeich F., Schaefer W., Baumeister W.;
RT "Nucleotide sequence analysis of the gene encoding the Deinococcus
RT radiodurans surface protein, derived amino acid sequence, and
RT complementary protein chemical studies.";
RL J. Bacteriol. 169:5216-5223(1987).
CC -!- FUNCTION: SHAPE MAINTENANCE. POSSIBLE PROTECTION FROM NOXIOUS
CC ENZYMES OR EXOGENOUS AND UNSETTLING DNA, AND MAY MEDIATE HOMOTYPIC
CC CELL-CELL CONTACTS.
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC HEXAGONAL S-LAYER.
CC -!- PTM: CONTAINS TIGHTLY BOUND REDUCING SUGARS (SIX PER POLYPEPTIDE
CC CHAIN) AND FATTY ACIDS (COVALENTLY BOUND AND LOCATED IN THE N-
CC TERMINAL REGION).
CC -!- PTM: THE N-TERMINUS IS BLOCKED.
CC -!- MISCELLANEOUS: THE HYDROPHILIC C-TERMINAL REGION RICH IN AROMATIC
CC AA COULD BE ENGAGED IN INTERACTIONS WITH NUCLEIC ACIDS, AND THE
CC BOUND FATTY ACIDS AND THE N-TERMINAL REGION COULD SERVE TO ANCHOR
CC THE LAYER TO THE OUTER MEMBRANE OF D.RADIOURANS. HPI LAYER
CC CONTAIN ABOUT 30% BETA STRUCTURE AND VIRTUALLY NO ALPHA HELIX.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL; M17895; AAA3335.1; --
DR PIR; A29832; A29832.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; UNKNOWN 1.
KW Glycoprotein; Lipoprotein; Signal; Cell wall; S-layer.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 1036 HEXAGONALLY PACKED INTERMEDIATE-LAYER
FT SURFACE PROTEIN.
FT DISULFID 74 86
FT DISULFID 256 275 PROBABLE.
FT DISULFID 642 754
FT DOMAIN 60 250
FT DOMAIN 985 1030 SER/THR-RICH (29%).
FT SEQUENCE 1036 AA; 108028 MW; AAF98206A74AEE CRC64;
QY 16 SSSHVNAKLGAROTRTPLLT-NFPGQA-PLPLTLQHTTE-SGANDTILPLNNIQDIL 72
Query Match 3.9%; Score 102.5; DB 1; Length 1036;
Best Local Similarity 20.2%; Pred. No. 8.2;
Matches 100; Conservative 61; Mismatches 181; Indels 153; Gaps 26;
```

```
Db 101 STSYTATATDAANKVCTSSVTVNVAGVSNPAPTTAVTLTDLAGVSSAPITIKDANGNV 160
QY 73 VGMKKOKERFVQVNDATSFKTKTYV-----PORTSAA 109
Db 161 QGYDN-----VTVDNATITVARGVTVTAGNVSGFNGPTTNRVLDLGGNQTVTLAY 213
QY 110 ILLISDSQOPLAFVNL-----GFSNTGLQALGIT-DDLGAQPPDGQFADAANLGGDLQ 163
Db 214 TOAGTTTPTPVGSGINILTPAVGTSVTGGSTVRVTFPKANEVQCMVGGAAVTAQVDSTG 273
QY 164 W-----VAPFTGTTHGVFLIGSDQDDFLDQFDDISSTFGSSITQVQALSGSGARPPDQAG 219
Db 274 YCDVVVPNSTGVV--ITVMGKGVNG-----QTVTATRNISVTOAAVSGVVTP---AG 322
QY 220 HEHFGL-DCISQPSVTGWETTVFPGQAVVPPGIIILTRDGDGTGRPSWALDGSFMAFRH 278
Db 323 DQELTLTSEGIVRADSGWRRL---CGGVSTP-----SDPNLNLDIYIKGTVN 367
QY 279 FQ-----QKVPEFNAYTLANAIPAN---SAG-----NLTOQEGAEPLGA 314
Db 368 FSVNAPAGQKVELFLARTTGSVDPTNDDIOAGDVLSVASTGTETFSLSRRLAEFDGV 427
QY 315 RMFGRW---KSGAPIDLAPTADD-----PALGADPORN-----NNFDYSDTLT 354
Db 428 R---KXIVVRINGTQVYQPVADNKGPOP---DPELNGVQVAYSNILNNYNSGL--- 478
QY 355 DETRCPFGAHVRKNRPNQDLGGPVDTFHAMRSSIPYGPETSDAEL-----ASGVTAQDR 408
Db 479 -----TVVR-----GPNVFTS-----NPSLQDREFGQAPVGSFVQRRPA 514
QY 409 GLLFVEYQSIIGNGF 423
Db 515 GFESIRYILVPESAF 529

RESULT 15
PANI_YEAST STANDARD; PRT; 1480 AA.
ID PANI_YEAST
AC P32521;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE PANI protein.
GN PANI OR MIP3 OR MDP3 OR YIR006C OR YIB6C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 320-344; 352-375 AND 899-906.
RX MEDLINE=92405166; PubMed=1339314;
RA Sachs A.B., Deardorff J.A.;
RT "Translation initiation requires the PAB-dependent poly(A)
RT ribonuclease in yeast.";
RT Cell 70:961-973(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copesey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Hornselli T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RX MEDLINE=95282515; PubMed=7762303;
RA Voss H., Tamames J., Teodoru C., Valencia A., Senses C., Wiemann S.,
RA Schwager C., Zimmermann J., Sander C., Ansoorge W.;
RT "Nucleotide sequence and analysis of the centromeric region of yeast
RT chromosome IX.";
```



```

RL Yeast 11:61-78(1995).
CC -1- FUNCTION: NOT KNOWN.
CC -1- PTM: THE N-TERMINUS IS BLOCKED.
CC -1- SIMILARITY: CONTAINS 2 EH DOMAINS.
CC -1- CAUTION: WAS ORIGINALLY THOUGHT TO BE A SUBUNIT OF PAB-DEPENDENT
CC POLY(A)-SPECIFIC RIBONUCLEASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@1sb-sib.ch).
CC -----
DR EMBL: Z38062; CAAB6208.1; -.
DR EMBL: M90688; AAA34841.1; -.
DR EMBL: X79743; CAB38097.1; -.
DR PIR: S30889; S30889.
DR PIR: S48440; S48440.
DR SGD: S0001445; PAN1.
DR InterPro: IPR000261; EPRS15_repeat.
DR Pfam: PF00036; ehband; 3.
DR SMART: SM00054; Efh; 2.
DR SMART: SM00027; EH; 2.
DR PROSITE: PSS0031; EH; 2.
DR Repeat.
KW DOMAIN.
FT 142 575 15 X 12 AA APPROXIMATE REPEATS.
FT REPEAT 142 153 1-1.
FT REPEAT 164 175 1-2.
FT REPEAT 188 199 1-3.
FT REPEAT 215 226 1-4.
FT REPEAT 235 246 1-5.
FT DOMAIN 270 359 EH 1.
FT REPEAT 392 403 1-6.
FT REPEAT 409 420 1-7.
FT REPEAT 422 433 1-8.
FT REPEAT 446 457 1-9.
FT REPEAT 467 478 1-10.
FT REPEAT 498 509 1-11.
FT REPEAT 507 518 1-12.
FT REPEAT 538 549 1-13.
FT REPEAT 545 556 1-14.
FT REPEAT 564 575 1-15.
FT DOMAIN 328 680 2 X 23 AA APPROXIMATE REPEATS.
FT REPEAT 328 350 2-1.
FT DOMAIN 600 689 EH 2.
FT REPEAT 658 680 2-2.
FT DOMAIN 1084 1125 7 X 6 AA TANDEM REPEATS.
FT REPEAT 1084 1089 3-1.
FT REPEAT 1090 1095 3-2.
FT REPEAT 1096 1101 3-3.
FT REPEAT 1102 1107 3-4.
FT REPEAT 1114 1119 3-5.
FT REPEAT 1120 1125 3-6.
FT DOMAIN 1315 1377 8 X 6 AA APPROXIMATE REPEATS.
FT REPEAT 1315 1320 4-1.
FT REPEAT 1321 1326 4-2.
FT REPEAT 1327 1332 4-3.
FT REPEAT 1340 1345 4-4.
FT REPEAT 1345 1350 4-5.
FT REPEAT 1355 1360 4-6.
FT REPEAT 1361 1366 4-7.
FT REPEAT 1372 1377 4-8.
FT DOMAIN 13 22 POLY-GLN.
FT DOMAIN 29 34 POLY-GLN.
FT DOMAIN 98 106 POLY-GLN.
FT DOMAIN 1400 1406 POLY-GLU.
FT DOMAIN 1452 1455 POLY-PRO.
FT DOMAIN 1474 1480 POLY-PRO.
FT CONFLICT 235 235 P -> T (IN REF. 1).
FT CONFLICT 266 273 ITADQAK -> YYCPRSGKN (IN REF. 1).

```

```

FT 474 487 MISSING (IN REF. 1).
FT CONFLICT 653 657 MISSING (IN REF. 1).
FT CONFLICT 1291 1291 A -> R (IN REF. 1).
FT CONFLICT 1396 1480 GGVLPPPLPPTPOASTSEPIIAHVNDYNGAEKGTAGYS
FT DSDDDVLSTPESVGTDEEAGQPVSTAGIPSPGIGPP
FT PPPLP -> EACGLHPHLIQANKLPQNLISLTITLTYWL
FT KARAHMDPIIMMTTFYRFNLQVLVOMKRRKGHNOFLVSH
FT QPHLOVFLHPHFHEDLICFL (IN REF. 1).
SQ SEQUENCE 1480 AA; 160267 MW; F518495FF759553 CRC64;

Query Match 3 9%; Score 101.5; DB 1; length 1480;
Best Local Similarity 21.8%; Pred.No.16;
Matches 100; Conservative 54; Mismatches 176; Indels 129; Gaps 24;

QY 34 PLLTNF-----PGQAPPLTLTQHTT-----ESGANDTILPLNNIQGDILVGMKKQKEREVFFQ 86
Db 218 PQQTGFYLPQONAPLEPLKPTATGPNSTAN-----NOLNNDI-----KIPAIRLSFIT 267
QY 87 VNDATSEFKALKTYVPQRITSAAILISDPSQPLAFVNLGFSNTGLQ--ALGITDLDGD 143
Db 268 AQDQAKFETLFRSIV-----TNGSNTVSGANCRKI-LMRGLPPSQLARIWTLCDTSKAGE 322
QY 144 AQPPDQGFADANLGDGLSGWVAPFTGTTIHGVFLIGSDDDDFLDQFTDIISSITFGSSIT 203
Db 323 LTFPP--EFALAMHLINDVLQ-----GDTI--PYELDSKTKNEVSSFIDAINDLSIANQ-- 370
QY 204 QVALSGSARPPDQ--AGHEHFGFLDGIQPSVTGWETTVPQQAVVPPGIIUTGRGD 260
Db 371 DSGANDAPKTPFDEFITAGVQNL-----QPOFTGMPOTSG--IFLQSOITGGGVA 420
QY 261 TGRPSWALDGSFMAFRHFOOKVBEFNAYTLANAIPANSAGNLTOEGAEFLGARMGFRW 320
Db 421 SALNPG--STGFMAPTTF-----NMSMTGTGLNPQIG-- 453
QY 321 KSGAPIDLAPTDPAI-----GADPQRNNNFYSDTLTETCPFGAVRKTNPRDGLG 375
Db 454 --GAPASMQPNITGNALQPOTTGMPPTTGMPQTTGMPOT--SFGVNL----- 499
QY 376 GPVDTFHAMRSSIPYGPETSDAELASGVTAQDRGLLFEVYSIINGNFRFOQIMANNAN 435
Db 500 -----GQLTGALOSQYTG-----GYGSVMP-----QDSGPASMPN 531
QY 436 PPSKRPITPGIEPIIGQTPRTVGGLDPLNQNETFTVPL 474
Db 532 LSPNQ--QGIGSQLTGLQDQPGTGFPP--SNFSATMPL 565

```

Search completed: November 27, 2002, 13:46:02
 Job time : 24 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 27, 2002, 13:45:31 / Search time 35 Seconds
(without alignments)
2931.756 Million cell updates/sec

Title: US-09-926-084-7
Perfect score: 2601
Sequence: 1 MDLSLFVSVAVLVGSSSHV.....KGEYFPLPSISALTATIAA 498

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP_TREMBL_21:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protist:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirts:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB ID	Description
1	2581	99.2	498	3	Q8WZK8
2	1396.5	53.7	499	3	P87212
3	412.5	15.9	460	16	Q9R208
4	386.5	14.9	469	16	Q8YWM0
5	185	7.1	421	16	Q9JUN5
6	177	6.8	421	16	Q9KJP5
7	143.5	5.5	434	16	Q8ZP62
8	133.5	5.1	423	16	Q8XAS4
9	128	4.9	1672	16	Q8Y366
10	126.5	4.9	1067	16	Q9R203
11	121.5	4.7	299	16	Q8ZCG6
12	119.5	4.6	1093	16	Q9CGV6
13	118.5	4.6	420	16	Q9RKO2
14	118	4.5	2468	16	Q912M3
15	118	4.5	2747	2	Q9L800
16	115	4.4	3972	16	P73139

17	113	4.3	1435	15	Q92898	Q92898 human immun
18	112.5	4.3	1248	10	Q9SAJ2	Q9SAJ2 arabidopsis
19	112.5	4.3	1832	5	Q96503	Q96503 cryptospori
20	111.5	4.3	523	4	Q9H9X1	Q9H9X1 homo sapien
21	111.5	4.3	984	4	Q96HA1	Q96HA1 homo sapien
22	111.5	4.3	2712	10	Q9SB74	Q9SB74 arabidopsis
23	110	4.2	782	2	Q93SH4	Q93SH4 bradyrhizob
24	110	4.2	4199	16	P74440	P74440 synechocyst
25	109.5	4.2	421	16	Q92ES2	Q92ES2 listeria in
26	109.5	4.2	421	16	Q8YA00	Q8YA00 listeria mo
27	109.5	4.2	1688	2	Q8R060	Q8R060 actinobacil
28	109.5	4.2	1742	16	Q55583	Q55583 synechocyst
29	109.5	4.2	4928	2	Q9ALM3	Q9ALM3 saccharopol
30	109	4.2	1976	2	Q9WX18	Q9WX18 bacillus sp
31	108.5	4.2	1003	15	Q90619	Q90619 human immun
32	108.5	4.2	3029	16	Q55582	Q55582 synechocyst
33	108	4.2	3778	12	Q84509	Q84509 parametium
34	107.5	4.1	2523	16	Q53393	Q53393 mycobacteri
35	107	4.1	642	5	Q9NH51	Q9NH51 drosophila
36	107	4.1	649	5	Q9SSS7	Q9SSS7 drosophila
37	107	4.1	940	16	Q8Y4N9	Q8Y4N9 listeria mo
38	107	4.1	1003	15	Q9YV12	Q9YV12 human immun
39	107	4.1	1606	12	Q91LD6	Q91LD6 white spot
40	106.5	4.1	1003	15	Q9WPY5	Q9WPY5 human immun
41	106.5	4.1	1008	15	Q97059	Q97059 human immun
42	106.5	4.1	1430	15	Q90VT5	Q90VT5 human immun
43	106.5	4.1	1433	15	Q9DH53	Q9DH53 human immun
44	106.5	4.1	1433	15	Q9QNM2	Q9QNM2 human immun
45	106	4.1	444	2	Q93HS1	Q93HS1 bradyrhizob

ALIGNMENTS

RESULT 1

Q8WZK8 ID Q8WZK8 PRELIMINARY; PRT; 498 AA.

AC Q8WZK8; 01-MAR-2002 (TREMBL:rel. 20, Created)
DT 01-MAR-2002 (TREMBL:rel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBL:rel. 21, Last annotation update)

DE DYP.
GN Geotrichum candidum (Oospora lactis).
OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Galactomycetes.
OX NCBI_TaxID=27317;
RN [1]
RP SEQUENCE FROM N.A.

RA Sugano Y., Sasaki K., Shoda M.;
RT Cloning and sequence analysis of a novel decolorizing enzyme,
RT peroxidase gene dyp from Geotrichum candidum Dec 1, 1998;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB013135; BAA77283.1; -
DR InterPro: IPR000169; SHProl acsite.
DR PROSITE; PS00639; THIOI-PROTEASE_HIS; UNKNOWN 1.
SQ SEQUENCE 498 AA; 53307 MW; 11BB3BC945B30DB7 CRC64;

Query Match 99.2%; Score 2581; DB 3; Length 498;
Best Local Similarity 99.4%; Pred. No. 1.6e-184;
Matches 495; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy	1	MDLSLFVSVAVLVGSSSHVNAALGARQRTTTPULNFPQOAPLPTLTQHTTSGANDT	60
Db	1	MDLSLFVSVAVLVGSSSHVNAALGARQRTTTPULNFPQOAPLPTLTQHTTSGANDT	60
Qy	61	ILPLNNTIGDILVGMKKQKRFVFOVNDATSPFTALKTVPORITSAIILSPSQOPL	120
Db	61	ILPLNNTIGDILVGMKKQKRFVFOVNDATSPFTALKTVPORITSAIILSPSQOPL	120
Qy	121	AFVNLGFSNTGIALGITDDGDAQPDGQPADANALGDDLSCQWVAPFTGTTIHGVFLIG	180
Db	121	AFVNLGFSNTGIALGITDDGDAQPDGQPADANALGDDLSCQWVAPFTGTTIHGVFLIG	180

Db 121 AFVNLGFSNTGLQALGITDLDGDAQFDGQFADAAANLGGDLDSQWVAPFTGTTIHGVFLIG 180
 QY 181 SDODDFLDQFTDDISSTFGSSITQVQALSGSARPPDOAGHEHFGFLDGSQPSVTGWETT 240
 Db 181 SDODDFLDQFTDDISSTFGSSITQVQALSGSARPPDOAGHEHFGFLDGSQPSVTGWETT 240
 QY 241 VFGQAVVPGIILTCRGDGTGRPSWALDGSFMAFRHFOOKYKPEFNAYTLANAIPANSA 300
 Db 241 VFGQAVVPGIILTCRGDGTGRPSWALDGSFMAFRHFOOKYKPEFNAYTLANAIPANSA 300
 QY 301 GNLTOEGAEFLGARFGRWKSGAPIDLAPTADDPALGADPQNNNPFDSYDITLDTETRCP 360
 Db 301 GNLTOEGAEFLGARFGRWKSGAPIDLAPTADDPALGADPQNNNPFDSYDITLDTETRCP 360
 QY 361 FGAHVKTNPQDLGGPVDTFHAMRSSIPYGPETSDAELASGVTAQDRGLLFFVEYQSIIG 420
 Db 361 FGAHVKTNPQDLGGPVDTFHAMRSSIPYGPETSDAELASGVTAQDRGLLFFVEYQSIIG 420
 QY 421 NGRFQOINWANNANFPFSPKPIPTGIEPIIGQTPRTVGGDLPLNQNETFTVPLFVPIKG 480
 Db 421 NGRFQOINWANNANFPFSPKPIPTGIEPIIGQTPRTVGGDLPLNQNETFTVPLFVPIKG 480
 QY 481 GEYFFLPSISALTATIAA 498
 Db 481 GEYFFLPSISALTATIAA 498

RESULT 2
 P87212 PRELIMINARY; PRT; 499 AA.
 AC P87212; (TREMBlrel. 04, Created)
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Peroxidase.
 GN CP021
 OS Polyporaceae sp.
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
 OC Aphyllophorales; Polyporaceae.
 OX NCBI_TaxID=54443;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Han Y.-H.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U77073; AAB58908.1; -
 DR InterPro; IPR000504; RNA rec mot.
 DR PROSITE; PS00030; RRM RNF 1; UNKNOWN 1.
 SQ SEQUENCE 499 AA; 53973 MW; EFBDA18EB6B8624 CRC64;

Query Match 53.7%; Score 1396.5; DB 3; Length 499;
 Best Local Similarity 55.9%; Pred. No. 4.8e-96;
 Matches 283; Conservative 63; Mismatches 145; Indels 15; Gaps 9;

QY 1 MDLSLFVSVAVLVGSSSHVNAKLGAROTRTTLLTNFPGQAPLTLTOHTTESGANDT 60
 Db 1 MQUKHFLAATAAP--SAVTOSAFAYHVKRARSPLTGSFPGPPLPTIAEVQSSAGNDS 58
 QY 61 ILPLNNIQGDLVGMKKOKERFVFFVNDATSEKTKALTYVVPQRTISAALISDPSQOPL 120
 Db 59 -LPFENIQGDLVGMKKOKERFVFFHINNATTFKSLKYAPANITSVQTIIGPASQOPQ 117
 QY 121 AFVNLGFSNTGLQALGITDLDGDAQFDGQFADAAANLGGDLDSQWVAPFTGTTIHGVFLIG 180
 Db 118 AFVNLAFSHTGFGALGVADDLQDTAFTAGQFADAPSGDDTSTWEAFKGTNDGVFLIG 177
 QY 181 SDODDFLDQFTDDISSTFGSSITQVQALSGSARPPDOAGHEHFGFLDGSQPSVTGWETT 240
 Db 178 SDVDTTNNQYRDDLKKLGDWVLLDLSAARPGAEKGEHFGFLDGSINPTIPGFGTP 237
 QY 241 VFGQAVVPGIILTCRGDGTGRPSWALDGSFMAFRHFOOKYKPEFNAYTLANAIPANSA 300
 Db 238 -FPGQAVVDSGVIFAGRTNDPVTNRRFSWALDGSFLVPRKLQVLPEFHKKWTLNAL-QNQ 295

QY 300 AGNLTOEGAEFLGARFGRWKSGAPIDLAPTADDPALGADPQNNNPFY-----SPTLTD 355
 Db 296 AGNLTVBEGALLGSRMFGWNSGAPIDLTDPVDDTLGNDPQNNNPFNIHPGEDLTDD 355
 QY 356 ETRCPFGAIVKRNTPQDL---GGPVDTFHAMRSSIPYGPETSDAELASGVTAQDRGLLFF 412
 Db 356 ETRCPFTAHVKTNP-R-DLEAQGLIPDLFHAIRAGTPYGPVTDAESNNTTISIDRGLAF 414
 QY 413 VEYQSIINGFRPQOINWANNANFPFSPKPIPTGIEPIIGQTPRTVGGDLPLNQNETFTV 472
 Db 415 VEYQSVISNGFRPQOINWANNANFPENKSEPLGLDVPVIGQGTQTF-GLDPRNASDLSLT 473
 QY 473 PLFVPIKGGYFELPSISALTATIAA 498
 Db 474 PQIISNGGEYFSPSITALVHEFGA 499

RESULT 3
 Q9R208 PRELIMINARY; PRT; 460 AA.
 AC Q9R208;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE Peroxidase, putative.
 GN DRA0145.
 OS Deinococcus radiodurans.
 OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
 OC Deinococcaceae; Deinococcus.
 OX NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1;
 RX MEDLINE=20036896; PubMed=10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome sequence of the radioresistant bacterium Deinococcus
 RT radiodurans R1.";
 RL Science 286:1571-1577(1999).
 DR EMBL; AE001862; AAP12218.1; -
 DR TIGR; DRA0145; -
 KW Peroxidase; Complete proteome.
 SQ SEQUENCE 460 AA; 51657 MW; CB8C476C1DD1AB50 CRC64;

Query Match 15.9%; Score 412.5; DB 16; Length 460;
 Best Local Similarity 29.3%; Pred. No. 1.3e-22;
 Matches 148; Conservative 68; Mismatches 190; Indels 99; Gaps 25;

QY 34 PLLTNPPGQAPLPTLTQHTTESGANDTI-LPLNNIQGDLVGMKKOKER-FVFPQVNDAT 91
 Db 2 PENTLF---KRLRELVHH-----NDKIDLDLDIQATVL-----RERPEPYGTHAMV 46
 QY 92 SFKTA-----LKYVVPQRTISAALISDPSQOPLAFVNLGFSNTGLQALGITDLDGDAQ 145
 Db 47 RDTAEGGRELKRLLP-HIASA-----EKWMDVKYANTAAAISEGLKKGVPQDSLDS- 100
 QY 146 FPD---GQFADAANLGD---DLQWVAPFTGTTIHGVFLIGSDOOD---FLDQFTDD 193
 Db 101 FPESFKVGMAGRAEHLFDVGENDPKHWEKPGFTGVHLLATTFIAENEENQKALVIAEHE 160
 QY 194 ISSTFGSSITQVQALSGSARPPDOAGHEHFGFLDGSQPSVTGWETTVEFGQA-VVPRGI 252
 Db 161 LEATKG--VTLMLREDFGAQP---DSRNLGYKDMISNPAIEGSGIKPPFGQGPALKPGE 215
 QY 253 ILTGRDGTGT-----RPS-WALDGSFMAFRHFOOKYKPEFNAYTLANAIPANSAGNLTOQ 306
 Db 216 FVLGYGPGAGVPLGMPKPEVLGKNGTFFVALRYHTNAGSFNRYLKENA--EYTGGD--- 269

```

QY 307 EGAEFLGARNFGKSGAPIDLAFTADDPALGADPQRNNNFEDYSOTLTDETRCPFGAHR 366
D 270 --AELLAAKLIVGRWRSGLPLTAPKEDDPELGHDPPNRNNDETYKND--PEGLEVPLGSHIR 326
QY 367 KTRNRO--DLGGEVDTFHAMRSSIPYGPETSDELAASGVTAAQDRGLLFVYOSIIGNGF 423
D 327 RMRPRDRLKLELTDVNIHRIIRATAYGPAYDPKADSLAEDKVERGYLFIFISKAMDTT 386
QY 424 RFOQINANNANFPFSKPTPTGIEPIIGQTPRTVGLDPL---NONETFTV----- 473
D 387 EFLQKEMINKANF-----IGQGSER-----DPIVGLODEDLTTLKEPVRQ 428
QY 474 -----LFVLPKGEYFPLPSISAL 492
D 429 RLRGMDTFENVLRGGEYLFPPLSLAL 453

RESULT 4
OBYMMO PRELIMINARY; PRT; 469 AA.
AC OBYMMO;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DE Hypochemical protein A1r1585.
GN A1r1585.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; Pubmed=11759840;
RA Kaneo T., Nakamura Y., Wolk C.P., Kunitz T., Sasamoto S.,
RA Kishida Y., Kohara M., Matsunoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Rep. 8:205-213(2001).
DR EMBL; AP003586; BAB77951.1; -.
KW Hypochemical protein; Complete proteome.
SQ SEQUENCE 469 AA; 53500 MW; 65328698C3C8D4C1 CRC64;

Query Match 14.9%; Score 386.5; DB 16; Length 469;
Best Local Similarity 29.9%; Pred. No. 1.2e-20;
Matches 144; Conservative 57; Mismatches 185; Indels 95; Gaps 23;

QY 64 LNNIQQDILIVGMKKQKERFV--FQVNDATSFKTALKTYYVQRTISA-----AIIISDP 115
D 28 LNDIQNLIKHGHGHDHSHLFLQFKRBEVVEVKKWIGSFAOTYITSAKKODEAFKTRQK 87
QY 116 SQQPLAFVNLGFSNTGLQALGIT--DLIGDAQFPDGFAD--AANLGD-DLSQWVAPFTG 170
D 88 GVSQDVAFNPLSHHGYEYLEIEPFOIPGDKPRFMGMKNEIRSLGDPKILATWELGFO- 146
QY 171 TTITGVPLISGDQDDFLQFDDISSTFGSSITQVQALSSGARPP--DQGH--EHNGF 225
D 147 SEIALVLIADDDIVDLQIVNOIT---OKLRQIAELVHREDGFIILNQOGLIIEHGF 202
QY 226 LDGTSQSVTQMETTVFPGAQV---PGIILITGSDGTGTPSPWAL 269
D 203 VDSQSP-----LFFKRDVNERVNNCPDKWDPKAPLDSILV--EDPKNGTQDST-- 251
QY 270 DGSFMAFHFQOKYVEFNAYTLANAIPANSAGNLTQEGAEFLGARMFGKSGAPIDLA 329
D 252 -GSYLVRKLEQNVYAFREDQKLAQKINIQENLA-----GALLVGRFADGTPVTL- 301
QY 330 PTADDPALGADPQRNNNFEDYSOTLTDETRCPFGAHRKTRNRODLDGGEVDTF----- 380
D 302 --SDPTVAVTP--TNNENYOGDLA--ATKCPHSHTRKTRNPGDTPARLLLTGDFDEAFK 356
QY 381 ----FHARRSSIPYGPETSDELAASGVTAAQDRGLLFVEYOSIIGNGFQOINMANNF 436

```

```

D 357 EERGHRIITRAVSVSGENNPESKEPVSG-----SGLLFLCFQSNIENQNFQMSRANQNF 411
QY 437 PFSKPTPTGIEPIIGQTPRTVGLDPLNQNTFTVPLFVIRKGEYFPLPSISA 491
D 412 ---VQVNTGPDPLIGPDSGTQKWPKKWG--EP--ETEEYNFQIWINNKGBYFPAPISF 464
QY 492 L 492
D 465 L 465

RESULT 5
OBYMMO PRELIMINARY; PRT; 421 AA.
AC OBYMMO;
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DE Possible periplasmic protein.
GN MMA0282.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65639;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; Pubmed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Davis K., Felwell T., Hamlin N., Holroyd S.,
RA Jørgensen M., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491.";
RL Nature 404:502-506(2000).
DR EMBL; AL162752; CAB83589.1; -.
KW Complete proteome.
SQ SEQUENCE 421 AA; 46119 MW; A598BF5DF42F562 CRC64;

Query Match 7.1%; Score 185; DB 16; Length 421;
Best Local Similarity 23.2%; Pred. No. 1.1e-05;
Matches 88; Conservative 41; Mismatches 125; Indels 126; Gaps 17;

QY 135 LGITDLDGDAQFPDGFQRA-----DAAN-----LQDDISQWVAPFTGTTIH 174
D 130 LTVTVGVSSSLF--DGRGLKDKRPILHLOEMRDFNSDKLQKSCDGLSLQICAFPTETCQ 188
QY 175 GVFLIGSDQDDFLQFDDISSTFGSSITO--VQALSGSARPPQAGHENGFLDGISQF 232
D 189 AA-----LNDIINKHTVQAVIRKSIDGWPQKSEGGMAARNLGFRDGTGPN 235
QY 233 SVTQMETTVFPGAQVVPBGILITGRDGTGTPSPWALDGSFMAFRHFOOKYVEFNAYTLA 292
D 236 KVSQDPKTA-----DEVLMTGVAANSLSDEREMAKNGSYQAVRLIRHVFEPDRPL- 285
QY 293 NATPANSAGNLTQEGAEFLGARMFGKSGAPIDLAFTADDPALGADPQRNNNFEDYSOT 352
D 286 -----OEQDIDIFGRKRY-----SGAPMDGKKEADQOPFADKPEGNIT- 322
QY 353 LTDETRCPFGAHRKTRNRODLSGP--VDTFHARRSSIPYGPETSDELAASGVTAAQDRGL 410
D 323 -----PKDSHIRLANFRD---PEFLKGRILFRRAVSY---SRGLASSG--QLDVLG 365
QY 411 LFVEYOSIIGNGFQOINMANNFPSKPTPTGIEPIIGQTPRTVGLDPLNQNTFTVPLFVIRKGEYFPLPSISA 470
D 366 VFVCCYQANLADGFIPLVQ--NLNG-----EPIEE----- 392
QY 471 TVPLFVLPKGEYFPLPSI 489
D 393 ----YISPFQGGYFFVLLPGV 408

```

```

RESULT 6
Q9K1P5 ID Q9K1P5 PRELIMINARY; PRT; 421 AA.
AC Q9K1P5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein NMB0036.
GN NMB0036.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Feden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.B., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Citton H., Clark E.B.,
RA Cotton M.D., Uterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizzo M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58.";
RL Science 287:1809-1815(2000).
DR EMBL; AF002362; AAF40507.1; -.
DR TIGR; NMB0036; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 421 AA; 46118 MW; 1488F54783074337 CRC64;

Query Match 6.8%; Score 177; DB 16; Length 421;
Best Local Similarity 23.4%; Pred. No. 4.5e-05;
Matches 89; Conservative 41; Mismatches 124; Indels 126; Gaps 18;

QY 135 LGITDLDGDAQFPDQFA-----DAAN-----LGDLSQWVAPFTGTTH 174
DB 130 LTVTVGVGSSLF-DGRFGLKDKKPIHLQEMRDFSNDKLOKSCWCDLSLQICAPTPTCQ 188
QY 175 GVFLGSDQDDFLDQFDDISSFTGSSITQ--VQALSGSARPFQDQAGHEHFGDGI 232
DB 189 AA-----LRDIKHTVQTAVIRWSIDGWPKSEPGMAARNLLGRDGTGPN 235
QY 233 SVTGWETTVPFGAVVPPGPIILTGRDQDGTTPRSWALDGSFMAFRHPQKVPBFNAYTLA 292
DB 236 KVSDPKTA-----DEVLTGTVANSILDEPEMAKNGSYQAVRLIRHFVFWDRTP 285
QY 293 NAIPANSAGNLTOEGAEFLGARMFGKSGAPIDLAPTADDPALGADPORNNNFVSDT 352
DB 286 -----QEQTDFGRKY-----SCAPMDGKKEADQDPFAKDPE----- 318
QY 353 LTBETRCFPAHVRKTNPRDLGCP--VDTFHAMRSSIPYCPETSDAELASGVTAQDRGL 410
DB 319 -GDIT--PKDSHIRLANPRD-----PEFLKKHRLFRAYS-----SRGLASSG--QLDVGL 365
QY 411 LFVEYQSIIGNFRFOINWANNANFPSPKIPITPGIEPIIGQTTPRTVGGDLPLNQNETF 470
DB 366 VFVCYQANLADGFIQVQ-NLLNG-----EPLLE----- 392
QY 471 TVPLFVLPKGEYFF-LPSI 489
DB 393 ----YISPFGGGVFFVLPGV 408

RESULT 7
Q82F62 ID Q82F62 PRELIMINARY; PRT; 434 AA.
AC Q82F62;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)

```

```

DE Hypothetical protein YPO1856.
GN YPO1856.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagers K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skellon J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
DR EMBL; AJ414150; CAC90673.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 434 AA; 47800 MW; AA3E8FB623D41B6 CRC64;

Query Match 5.5%; Score 143.5; DB 16; Length 434;
Best Local Similarity 22.9%; Pred. No. 0.015;
Matches 81; Conservative 38; Mismatches 123; Indels 111; Gaps 16;

QY 145 QPDPGQFADAANLGDLSQWVAPFTGTTHGVFLIGSDQDDFLDQFDDISSFTF---GSS 201
DB 173 RFPNDSLADGLCHGDVLMQLCANTNETVIHAL-----RDIKHTPDLLSVRKREGFI 225
QY 202 ITQVQALSSARPFQDQAGHEHFGDGIQSPSVTGWETTVPFGQAVVPPQIILTGRDGT 261
DB 226 SAHAARSKGQDTPINL-----LGFKDGKTANPKIS-----NKPLINNVVVSNNAGE- 271
QY 262 GTRPSNALDGSFMAFRHFQKVPEFNAYTLANAI PANSAGNLTOEGAEFLGARMFGRWK 321
DB 272 ---PAWAGVGSYQVRIIRPKV-EFWDRT-----PLOEQO-----TIFGRDK 309
QY 322 -SGAPIDLAPTADDPALGADPORNNNFVSDTLTDETRCPFGAHVRKTNPRDLGPGVD 380
DB 310 NSCAPLGMQHEHDEPNVAKDPECK-----VTPMDAHLRANPR-----TET 351
QY 381 FH--AMRSSIPYCPETSDAELASGVTAQDRGLLFVYQSIIGNFRFOQINWANNANFPF 438
DB 352 QRNLMLRRGYSYSLGVS-----SG--QLDMGLLFVCYQSDLAQAFLTVOERLNGALEEY 405
QY 439 SKPTPGIEPIIGQTTPRTVGGDLPLNQNETFTVPLFVLPKGEYFF-LPSIS 490
DB 406 VKPI-----GGGYFTTLPGVA 421

RESULT 8
Q8XAS4 ID Q8XAS4 PRELIMINARY; PRT; 423 AA.
AC Q8XAS4;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Orf, hypothetical protein.
GN YCDB OR Z1521 OR ECS1265.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,

```

```
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blatner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=2156231; PubMed=11258796;
RA Hayaishi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tohe T.,
RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
DR EMBL: AE005301; BAG55637.1; -
DR EMBL: AP002554; BAB4668.1; -
KM Complete proteome.
SQ SEQUENCE 423 AA; 46620 MW; 256FB987670B750C CRC64;

Query Match 5.1%; Score 133.5; DB 16; Length 423;
Best Local Similarity 22.3%; Pred. No. 0.081;
Matches 81; Conservative 29; Mismatches 119; Indels 135; Gaps 14;

QY 145 QPDDGQADANLGDLSQWVAPFTGTHGVFLGSDQDDFL-----DQFTDISSTF 198
DB 162 RFPDSDLAALCHGDVLLQICANTQDTVTHALRDIKHPTPLLSVRMKREGFISDHARS 221
QY 199 GSSITTVQVALSGSARPFQDQGHFGLDGISQSVTG-----WETTVFPGQAVVP 250
DB 222 KCKETPNNL-----GFKDGTAMPDSQNDKLMOKVWVTA----- 256
QY 251 GIITGRBDGTGRPSMALDGSFMAFRHFOQKVPFNAYTLANAI PANSAGNLTOEGAE 310
DB 257 -----DQEPRAWITGSSYQAVRLQPRV-FFWRPT-----PLKEQO--- 291
QY 311 FLGARMFGRMK-SCAPIDLAPTADDPALGADPQRNNNPDYSDDLTDTRCFGAHVKTN 369
DB 292 -----TIFGRDKOTGAPLQMHQEHDPDYASDPBKSG-----IALDSHRI LAN 334
QY 370 PROLDGPFVDTFHMRSSIPFGPETSDAELASGVTAQ--DRGLLFVYQSIIGNGFFQ 426
DB 335 PRT---AESSESLMLRRQSY-----SLGVNLSGOLMGLFVGYOHDLEKGFLLV 382
QY 427 QINMANNANFPFSKPTGIEPIIGTTPTTVGGLDPLNQNETHVPLFVLPKGEYFF- 485
DB 383 OKRLNGEALREYKFI-----GGGYFFA 405
QY 486 LPST 489
DB 406 LPGV 409

RESULT 9
Q8Y366 PRELIMINARY; PRT; 1672 AA.
AC Q8Y366;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Putative hemagglutinin-related protein.
GN RSC0115 OR R800985.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Beta subdivision; Ralstonia group;
CX Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=2161879; PubMed=11823852;
RA Satanoubat M., Genin S., Artiguenave F., Guzy J., Mangelot S.,
RA Ariat M., Billault A., Broctier P., Camus J.C., Catolico L.,
RA Chaudier M., Choiste N., Claudel-Renard C., Cunnac S., Demange N.,
```

```
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguler P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum."
RL Nature 415:497-502(2002).
DR EMBL: AL646057; CAD13643.1; -
KM Complete proteome.
SQ SEQUENCE 1672 AA; 164046 MW; 55058448B10FE1BB CRC64;

Query Match 4.9%; Score 128; DB 16; Length 1672;
Best Local Similarity 21.3%; Pred. No. 1.7; 214; Indels 106; Gaps 23;
Matches 105; Conservative 67; Mismatches 106; Gaps 23;

QY 30 TRTPPLTNFPGQAPLPTLTQHTTESGANDTIPLNNIQCGLVGMKKOKERFVFPQVND 89
DB 701 TRSYSLTVAAPTIALPTTLTAGATVSSA-----YQSV 733
QY 90 ATSPKTLAKTYVFORITSAALISDPQOPLAFVNLGFSNTGLALGITDDLDGAQFPDG 149
DB 734 AASGTAPTYVA---VTSGAL---PA-----GLSLSSAGV-----LSGTPTAGG 771
QY 150 QFADANLGDLSQWVAPFTGTHGVFLGSDQDDFLDQFTDISSTFGSITVOVALS 209
DB 772 SFSVTIS-ATDSTTGSQPFQGSRAV-TLVGSPPLTISPASTAGLTMAAGTSYQSFSAG 829
QY 210 GSARPFQDQGHFGLD-GIS-----QPSVTGWETTVFPGQAVVPGLI---LTGRDQDT 261
DB 830 GGVSPYTYALTIVNTGTPMAGLSFMAASATLSGTPTA-----GTGSTVTVATDSS 880
QY 262 GTRPSMALDGSFMAFRHFOQKVPFNA--YTLAN-AI PANSAGNLTOEG- AEFILGARMF 317
DB 881 GAGP-YAVSGTYT---LTVSAPTLTVAPATLPMPALGTAVSQSITASSGTAPTYAVTS 935
QY 318 GRMKSAPIDLAPTADDPALGADPQRNNNPDYSDDLTD-----RCFPGAHVKTN 369
DB 936 GALPAGLSLSA-----GVLSGTPTAGSFGFVTATDANSFTASRAYSITIGATVALN 990
QY 370 PROLDGPFVD-----TFHMRSSIPFGPETSDAELASGVTAQ--RGLLFVYQSIIGNGF 423
DB 991 PATVPGATLNTAVSQITFASGIGFPTTYAVASGTLBPAGVSLNSTGVLSGTPTLGSSTF 1050
QY 424 RFQOINMANNANFPFSKPTGIEPIIGTTPTTVGGLDPLNQNETHVPLFVLP---K 479
DB 1051 SIRATDSSTGAGAPYTG--TRGYTLVVQO-----AIGTAPPTATTTSTVPLPTANAT 1104
QY 480 GGEYFLPSTSA 491
DB 1105 GGPFSVTVIAA 1116

RESULT 10
Q9RZ03 PRELIMINARY; PRT; 1067 AA.
AC Q9RZ03;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Extracellular nuclease, putative.
GN DRB0067.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
CX Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Motil K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Ueberback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
```

RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL: AE001826; AAF12592.1; -.
DR TIGR: DRB0067; -.
DR InterPro: IPR005135; Exo_endo_phos.
DR Pfam: PF03372; Exo_endo_phos; 1.
KW Plasmid; Complete proteome.
SQ SEQUENCE 1067 AA; 109780 MW; D4D3B79B38F72B30 CRC64;

Query Match 4.9%; Score 127.5; DB 16; Length 1067;
Best Local Similarity 23.0%; Pred. No. 0.93;
Matches 121; Conservative 60; Mismatches 171; Indels 175; Gaps 28;

QY 13 LVGSSSHVNA---KLCAROTRTPL-----LNFPQAP 44
DB 55 LSGAGSAIKQASVQLGQAQATEVGLSFKNNGYGVYDSSNNLTMWATFDVNTSDQ-P 113
QY 45 L-PTUQTHTTEGANDTI--LPLNNIQDILVGMKKQKRFVFFQVNDATSF-----93
DB 114 LAAPTIPVDTE-GAGTICTTAFKNVR-----YFGSDASSRAPLLAFD 157
QY 94 -----KTALKTYVPQRTISAIIISDPSQOPLAFVNLGFSNTG--LOAL--GITDDLGDA 144
DB 158 SSNDPKTAVKTLRLDLDGSGVQNLPAQLA---GTSHTGKMGALPAGAT-----207
QY 145 QPDPGOFADANLGGDLDSQWVAPFTGTTIHGVFLIGSDQDDFLDQFDDISSTFGSSITQ 204
DB 208 -----GQVTLAASIPAAASNSQDNFSS-----FDLVFTADVNPSTLTINIGA 249
QY 205 VQALS-GSARPPDQAGHEHFGFLDGISQPSVTGNETTVFPQAVVPPGIILTG-----RDG 259
DB 250 VCGOTPGDRPALTAQ-----TIEGVVTAVEPG-----LSGFFVQEEG 289
QY 260 DTGTRSWALDGSFMAFRHFQKVPENAYTLANAIPANSAGNLTOEG--AEFLGARMF 317
DB 290 IDADRDETSDGLFV-----YCAASC-PALSAGDRVRVSGTVABYGGATQM 334
QY 318 -----GRWKSAPIDLAFTADDPALGADPQNNFND-----YSDTLTDTETRCPCFGAHRK 367
DB 335 TAPTVTKLLSGLA---PPAAELKLPDKTQERYEGMRVFPETLTITNNYTYGRY---388
QY 368 TNPQDLGGPVDTFHAMRSSIPYCPETSDAE-----LASGVTAQ---DRGLLFVYQSII 419
DB 389 -----GQLDSNAGRMFNPTNGASASEQSTITLDDGVSQNPDLNLYLSAERTRT 440
QY 420 GNGFRQOINWANNAPFSPKPTPGIEPIIGQTT-----PRTVGG 460
DB 441 GDTVTGLSGVMHVSANQPMLEP--EGAVEFVSANSRASNALPHDVGG 485

RESULT 11
Q8ZCG6 PRELIMINARY; PRT; 299 AA.
AC Q8ZCG6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein YPO3025.
GN YPO3025.
OS Versinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Versinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titchall R.W., Holden M.T.G.,
RA Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,

RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagers K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
DR EMBL: AJ414155; CAC92267.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 299 AA; 33389 MW; 5569FD4D38F8B8C3 CRC64;

Query Match 4.9%; Score 126.5; DB 16; Length 299;
Best Local Similarity 21.7%; Pred. No. 0.16;
Matches 91; Conservative 45; Mismatches 109; Indels 175; Gaps 23;

QY 109 ALIISDPSQOPLAFVNLGFSN--TGLQALGITDLDGDAQFPD---GQFADANLGGDL 162
DB 16 AIFMEAKVQGLDAIRLGCKKFCQSLQEL-----QQQFPDEHLGAVIAGSNVMDLS 68
QY 163 -----QWVAPFTGTTIHGVFLIGSDQDDFL---DQFTDDISSTFGSSITQVQALSGSARP 214
DB 69 NQGAKEKLPFVPL---GKGLAPATORDLLIHQSURQDINFTLAQAAV---AAGSATA 122
QY 215 FDAQH-----EHPGFLDGISQPSVTGNETTVFPQAVVPPGIILTGRDGTGTRPS 266
DB 123 VEETHGFRWVERDFTGIDGTENP-----QGDKEPEVAIADG-----BEDAG---167
QY 267 WALDGSFMAFRHFQ-----QKVPENAYTLANAIPANSAGNLTOEGAEFLGARMPGRW 320
DB 168 ---GSYVLVQRYEHNLNKQRIPE-----NEQE-----KIIGRT 198
QY 321 KSCAPIDLAFTADDPALGADPQNNFNDYSDTLTDTETRCPCFGAHRKTNPRQDLGGPVD 380
DB 199 K-----LDSQELPSD-QR-----PDTSHVSRVDLKENGK-----227
QY 381 FHAMRSSIPYCPETSDAEASGVTAQDRGLLFVEY-----QSTIGNGFRPQOINWAN 432
DB 228 LKILRSLPYG-----LASG-----KHGLYIAYCARLHNIEQQULS---MFGDIDGKH 273
QY 433 NANFPSPKPTPGIEPIIGQTTPTRTVGGDLPLNQNETFTVPLFVIPKGGEYFPLPSISAL 492
DB 274 DQLLRFSPKPT-----GSYVPAPSLTAL 296

RESULT 12
Q9CGV6 PRELIMINARY; PRT; 1093 AA.
AC Q9CGV6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein ykbc.
GN YKBC OR LLO986.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Mauger S., Jaillon O., Malarre K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403.";
RL Genome Res. 11:731-753(2001).
DR EMBL: AE006333; BAK05084.1; -.
DR InterPro: IPR000504; RNA_rec_mot.
DR PROSITE: PS00030; RRM_RNP_1; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1093 AA; 116310 MW; 4615B4393FB50317 CRC64;
Query Match 4.7%; Score 121.5; DB 16; Length 1093;
Best Local Similarity 21.3%; Pred. No. 2.7;

Matches 123; Conservative 64; Mismatches 228; Indels 163; Gaps 27;

QY 17 SSHVNAKLGAROTRTP-LITNPFQAPLPTLTQ-HTTESGANDTILPLNIGDILVG 74
 DB 62 SPVNSSLASDSTTEPPAVSNMENTAQSSTNITLSTESTYDNTMSPTOSONSIAIS 121
 QY 75 MKQKEREVEFQVNDATSPKALTYYVPRITSAAILISDPSQPLAFVNLGFSN----- 129
 DB 122 LTKASRESSINEPASTIANETTTADVTQTATQAQSDPTTS-LSESQKPTNPSQS 180
 QY 130 -----TGLQALGITDD--LGDQPFDDQFADANALGDDLQSWAPFTGTHGFLIGSD 182
 DB 181 KSTEITNIQVTVADVANNATGSAVF-----DGVNI-----TLQG----- 213
 QY 183 ODPFLDFTDISSTFSSITQVOALSGSARFPDQAGHEHFGLDGISOPEVTGNETTVF 242
 DB 214 -KDTTDLNLDPSGLHMSBOTQVIAIKGTAT--GQLNDENGVDDGPILPAT--YTMN 267
 QY 243 PGQA--VVPFGIILTGRGD-----TGTSPW-ALDGSFMAFHHFOQKVEFNAYTLAN 293
 DB 268 DGDGGRITVYVKSLSGLDLDMITVVASSDKDSQWQNEGA-----EGIPQGLTGEQ 319
 QY 294 AIPNANSKN--LTQDEGAEFLGAMFGKMSGAPIDLAFTAD-----DPLAGD----- 340
 DB 320 NI-AESGSGNSIVCLYNGANAL-SLIYQIVKHDTTEVPVVASFITTDIDNAQVQTNLAN 377
 QY 341 -----PQRNNFVSDTLTDET-----RCRPFQAHVR----- 366
 DB 378 LVTLPLPTNLKQGDGDTTYDASPNYPGLDGVASLYGGYLGAGVSEYYVYVAPABERA 437
 QY 367 -----KTNPRQDLGQVDTFHAMRSSI-----PYGPETSDAELASGVTAODRGL 410
 DB 438 DDSYFPAQVGRYDLFGSALQAH-WNTQIRQNFVYVYDEFGHKIQETHYVFGFGQDYNL 496
 QY 411 LFVEYQSIINGF-----RFOQIMANNANFP-----ESKPTTP 444
 DB 497 ---PIPIKGGFVNLTEENDASKNNPVINLIYHNHLPYYGNHNNNIYOGTAYTPSPFTI 553
 QY 445 GIEPIIGQTPRTVGLDPLAQNEFTVPLFVIRKGE 482
 DB 554 GYQNI---GNDEASITTYPVNNGKASVTLPLMAGK 588

RESULT 13

Q9RKQ2 PRELIMINARY; PRT; 420 AA.

AC Q9RKQ2;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Putative membrane protein.
 GN SC02276 OR SCC75A.22.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID:1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AS(2) / M145;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Klesner H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornby T., Howarth S.,
 RA Huang C.-H., Klesner I., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabbittowles E., Rajadream M.A., Rutherford K., Ruter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor AS(2).";
 RL Nature 417:141-147(2002).
 DR EMBL; AL133220; CAB61722.1; -
 SQ SEQUENCE 420 AA; 44522 MW; 56344615594E01BC CRC64;

Query Match 4.6%; Score 119.5; DB 16; Length 420;
 Best Local Similarity 21.9%; Pred. No. 0.89;
 Matches 81; Conservative 35; Mismatches 121; Indels 133; Gaps 17;

QY 89 DATSFKALTYYVPRITSAAILISDPSQPLAFVNLGFSNTGLQALGITDDLGDA---- 144
 DB 125 DDTEGALGLK--PSRL-----LRTGFGPSLFTTRGLADLRPEALADL 165
 QY 145 -QPPDGGFADANALGDDLQ--QWAPFTGTHGV---FLIGSDQDDFLDQFTDISSTFG 199
 DB 166 PKPF-GNULPRANSGLCVQACADDPQVAVHAIHNLARIG-----FG 207
 QY 200 SSITQVQAL-----SGSARFPDQAGHEHFGLDGISOQSVTGMETTVPPQAVVPPGIILT 255
 DB 208 KVVVRWSQLGFKGTSSTTPPEQTPRNILGFKDGRN-----IA 245
 QY 256 GRDGDGTGRPSWALD-----GSFMAFHHFOQKVEFNAYTLANAIANSAGNLTOQ 306
 DB 246 GTEKDRLDREVVMAEKDGTMTGSGSYLVARRIRMHIEIWD-----RASLQEQ 293
 QY 307 EGAEFLGARMFGRWK--SGAPIDLAFTADDPALGA-DQRNNFVSDTLTDETRCRPGAH 364
 DB 294 ED-----VFGHDKGSGAVGKAKERDEPFLAKMKPD-----AH 326
 QY 365 VRKTNPRQDLGQVDTFHAMRSSI-PYGPETSDAELASGVTAODRGLLFVEYQSIINGFR 424
 DB 327 VRLAHPSDNGATL-----LRGYSFTDGT-----DGLRLDAGLFFFLAYQDRIRIGFV 375
 QY 425 FQOQIMANNA 434
 DB 376 PVQRNLATDA 385

RESULT 14

Q912M3 PRELIMINARY; PRT; 2468 AA.

AC Q912M3;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Hypothetical protein PA1874.
 GN PA1874.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
 RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., LaRou M.,
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.B., Coulter S.N., Folger K.R., Kae A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Mong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen.";
 RL Nature 406:959-964(2000).
 DR EMBL; AE004613; AAC05263.1; -
 DR InterPro: IPR001343; Hemjysn_Ca_bind.
 DR InterPro: IPR003880; ppanne_attach.
 DR PRINTS; PR00313; CABNDNGRPT.
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 2468 AA; 238414 MW; 13596AFAB2C4B899 CRC64;

Query Match 4.6%; Score 118.5; DB 16; Length 2468;
 Best Local Similarity 21.1%; Pred. No. 16;
 Matches 115; Conservative 57; Mismatches 218; Indels 155; Gaps 28;

QY 21 NAKL-GARQRTTPLTLTNPFQAPLPTLTQHTTESGANDTILPLNIGDILVGKKKQ 79

```

Db 1088 NGSLSGTAPGTSVILTDNGN-----PIAEVTADSGNWTYPSTPIANGTVVNVVQAQ- 1142
QY 80 ERFVFOVNDATSEKALTALKTYVQORITSAAILISDPSCQQLAFVN--LGSNTGLQALGI 137
Db 1143 -----DASG-----NSSPPATVTVDDSSAPPAPVINPNSGVVISCTAEAGA 1182
QY 138 TDDLGA-QPDCOFADAANLGDLSQWAPFT-GTTI-HGVFLIGSDQDDFLDQFDDI 194
Db 1183 TVLTLDAGNPIGQV-----ADSGNW--SFTPGTPLANGTVIVAT-----ATDPT 1227
QY 195 SSTFGSITQVQALSGSARPEQAGHEHFGFLDGISOPSVTGMETTVFPGQAVVPPGIIIL 254
Db 1228 GNTGPOAATTVDAVAPAPVID-----PS-----NGTISGTAEGAKVIL 1268
QY 255 TGRDGDGTGRPSWALDGSFMAFRHQKVFEEFNAYTLANAIPANSAGNLTOQOGEAEL-- 312
Db 1269 T--DNGNPIGETTADGS--GNMSFTCTPLANG-TVVNAVAQDPAGN-TGPOGSTTVDA 1322
QY 313 -----GARMGRKWSGAPIDLAPTADDPALGADPQNNNFYSDTLTDETRCP 360
Db 1323 VAPNTPVNVPSNGLLNGTAEPGSTVTLTDGNGNPIGOTTADSGNWSF-----TPGSQLP 1378
QY 361 FGAHVKTNRQDLGG---PVDTFHAMRSSIPYGPET--SDAELASGVTAQDRLFLVE 414
Db 1379 NGTVNVV--ASDAAGNTSLPAT--TVDSLSIPQVDPNSGVISGTADAGNTIIITD 1434
QY 415 YQSIIGNGRFQOI-----NWANNANFPFS-----KP 441
Db 1435 -----GNGNPIGQVATADSGNWSFTPGIPLPDGTVVNVVARSPSNVDSAPAVITVDGVAP 1489
QY 442 ITPGIEPI-----ICQTPRTVGGLDPLNONTFTVPLFVIPKGGEYFPLPSISA 491
Db 1490 AAPVIDPSNCTEISGTAEAGATVILTDGNGNPIQ-----ATADSGNWTFTPTSTPL 1541
QY 492 LTATI 496
Db 1542 ANGTV 1546

```

RESULT 15

```

Q9L800 PRELIMINARY; PRT; 2747 AA.
ID Q9L800
AC Q9L800;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE RTX protein.
GN ASX.
OS Aeromonas salmonicida.
OC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
OC Aeromonas.
ON NCBI_TaxID=645;
RX (1)
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33656;
RA Braun M., Frey J., Kuhnert P.;
RT "280 kDa RTX protein of Aeromonas.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF218037; AAF27914.1;
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR001343; Hemlysn_Ca_bind.
DR InterPro; IPR003410; Hyalin.
DR InterPro; IPR002035; VWF A.
DR Pfam; PF00353; hemolysinCabin; 9.
DR PRINTS; PR00313; CABNDNGRPT.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE; PS00330; HEMOLYSIN_CALCIIUM; 4.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 2747 AA; 280201 MW; 208FE380E44A5F37 CRC64;

```

Query Match.

4.5%; Score 118; DB 2; Length 2747;

```

Best Local Similarity 19.7%; Pred. No. 20;
Matches 104; Conservative 57; Mismatches 204; Indels 162; Gaps 22;
QY 32 TTPLLTNFGQAPLPTLTQHTTESGANDTILPLNNIQGDIIVGMKKQKRFVFFQVNDAT 91
Db 288 TTPITITGTTDAASGSTVTLVVDANGNQOTLSAT-----VQPNG 326
QY 92 SFKTLKTYVQORITSAAILISDPSCQQLAFVNLSGNTGLQALGITDDLGAQFPDQGF 151
Db 327 TYSTDVTSPLPDGSDYDVTSVTDPAAGTGTATDDGSDVDVAPTTLAIT---LDANIITDDV 383
QY 152 ADAANLGDLLSQWVAPFTGT-----TIHG-----VFLIGSDQD 184
Db 384 INNAEAGQOI-----PVSQTVSGEFKAGDTVTLTVNGKPTGPVDADGRFTILVAGSD- 436
QY 185 DFLDQFTDDISSTFGSSITQVQALSGSARPEQAGHEHFGFLDGISOPSVTGMETTVFPG 244
Db 437 -----LAADTDRTIDASVTSTDAAGNSATATDSEGY-----GVDTTAPVD 476
QY 245 QAV-----VPGIILTG--RDGDTGT-----RPSWA-----LDGSPM 274
Db 477 LAITLDANITDDDDVINAAEAGQOIPIVSGTVTGEFKAGDTVTLTVNGKPTGPVDADGRFT 536
QY 275 AFRHFOOKVPEFNAYTLANAIPANSAGNLTOQOGEAFLGARMFGRWKSGAPIDLAPT--- 331
Db 537 ILVAGSOLAADTDRITDASVTSTDAAGNSATATDSEGYV-----DTTAPVDLAIITLDA 590
QY 332 --ADDPALGADPQNNNFYSDTLTDETRCPFGAHVKTNRQDLGGPVD-----TFHAM 384
Db 591 NITDDDDVINA-AEAGQOIPIVSGTVSGEFKA--GDTVTLTVNGKPTGPVDADGRFTILVA 647
QY 385 RSSIPVGPSTDAELASGVTAQDRLFLFVEYQSIIONGFRFOQINWANNANFPFS----- 439
Db 648 GSDL--AADSORTIDASVTSTDA-----GN-----STTKSDTENYTVDTVAPD 689
QY 440 -----KPITPGIEPIIGQT-----TPRTVGGI--DPLNQNETFTVPL 474
Db 690 LGIDLDPPIAVGGDNNVQAEADGKTPTVTLSGTVNGDANVGDVTLTL 736

```

Search completed: November 27, 2002, 13:47:49

Job time : 41 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: November 27, 2002, 13:45:31 ; Search time 16 Seconds
(without alignments)
915.788 Million cell updates/sec

Title: US-09-926-084-7
Perfect score: 2601
Sequence: 1 MDLSLFVSVAVLVGSSSHV.....KGEYFFLPISALTATIAA 498

Scoring table: BLOSUM62
Gap 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: Issued_Patents_AA:*
2: /cgn2_6/ptodata/1/iaa/5A.COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/5B.COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6A.COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/6B.COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	112.5	4.3	1837	US-08-928-361B-5	Sequence 5, Appl
2	109.5	4.2	4928	US-09-036-987A-5	Sequence 5, Appl
3	109.5	4.2	4928	US-09-370-700-5	Sequence 5, Appl
4	102	3.9	2089	US-08-418-893D-23	Sequence 23, Appl
5	102	3.9	2089	US-08-418-893D-24	Sequence 24, Appl
6	101.5	3.9	626	5268290-2	Patent No. 5268290
7	100	3.8	1003	US-07-743-357-10	Sequence 10, Appl
8	98.5	3.8	998	US-08-415-788-7	Sequence 7, Appl
9	98.5	3.8	1004	US-07-743-357-7	Sequence 7, Appl
10	98.5	3.8	1029	US-08-415-788-3	Sequence 3, Appl
11	97.5	3.7	913	US-07-743-357-22	Sequence 22, Appl
12	97.5	3.7	1005	US-07-743-357-1	Sequence 1, Appl
13	97	3.7	1003	US-07-743-357-9	Sequence 9, Appl
14	96	3.7	647	US-08-485-355B-50	Sequence 50, Appl
15	96	3.7	647	US-09-194-613-3	Sequence 3, Appl
16	96	3.7	675	US-08-485-355B-52	Sequence 52, Appl
17	95.5	3.7	1016	US-07-743-357-5	Sequence 5, Appl
18	95	3.7	490	US-09-253-149B-26	Sequence 26, Appl
19	95	3.7	699	US-08-694-865-16	Sequence 16, Appl
20	95	3.7	699	US-09-124-491-16	Sequence 16, Appl
21	95	3.7	1003	US-09-309-572-17	Sequence 17, Appl
22	94	3.6	1003	US-07-743-357-8	Sequence 8, Appl
23	93.5	3.6	414	US-09-715-834-2	Sequence 2, Appl
24	93.5	3.6	1381	US-09-540-245A-16	Sequence 16, Appl
25	92.5	3.6	501	US-08-913-477-23	Sequence 23, Appl
26	92.5	3.6	1016	US-07-743-357-3	Sequence 3, Appl
27	92.5	3.6	1042	US-08-928-361B-11	Sequence 11, Appl

28	91	3.5	316	1	US-07-828-980A-2	Sequence 2, Appl
29	91	3.5	1183	4	US-09-134-001C-3530	Sequence 3530, Ap
30	90	3.5	299	6	5514590-4	Patent No. 5514590
31	90	3.5	661	2	US-08-525-742-2	Sequence 2, Appl
32	89.5	3.4	353	2	US-08-687-702-37	Sequence 37, Appl
33	89.5	3.4	670	4	US-09-328-501-15	Sequence 15, Appl
34	89.5	3.4	1721	3	US-08-700-651-5	Sequence 3, Appl
35	89.5	3.4	1721	3	US-08-928-361B-6	Sequence 6, Appl
36	88.5	3.4	913	2	US-07-743-357-6	Sequence 6, Appl
37	88	3.4	513	4	US-09-071-035-68	Sequence 68, Appl
38	88	3.4	532	4	US-09-071-035-66	Sequence 66, Appl
39	88	3.4	746	5	PCT-US95-10509-2	Sequence 2, Appl
40	88	3.4	1140	4	US-09-220-081-2	Sequence 2, Appl
41	88	3.4	1140	4	US-09-677-575-2	Sequence 2, Appl
42	87.5	3.4	5215	4	US-09-105-537-2	Sequence 2, Appl
43	87	3.3	529	1	US-08-484-815-12	Sequence 12, Appl
44	87	3.3	529	3	US-08-888-949-12	Sequence 12, Appl
45	87	3.3	529	4	US-08-888-950-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-08-928-361B-5
Sequence 5, Application US/08928361B
Patent No. 6071518
GENERAL INFORMATION:
APPLICANT: Petersen, Carolyn
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
TITLE OF INVENTION: SPECIES INFECTIONS
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: PETERS, VERNY, JONES & BIKSA
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928,361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Verry, Hana
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.76-1 (HV)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1837 amino acids
TYPE: amino acid
STRANDEDNESS: linear
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-928-361B-5
Query Match 4.3%; Score 112.5; DB 3; Length 1837;
Best Local Similarity 18.5%; Pred. No. 0.12;
Matches 108; Conservative 64; Mismatches 168; Indels 245; Gaps 30;

LENGTH: 4928
 TYPE: PRT
 ORGANISM: Saccharopolyspora spinosa
 US-09-370-700-5

Query Match 4.2%; Score 109.5; DB 4; Length 4928;
 Best Local Similarity 23.5%; Pred. No. 1.2;
 Matches 100; Conservative 40; Mismatches 169; Indels 117; Gaps 24;

QY 30 TRTPPLTNFGQAPLPLTLTQHTTESGANDTILPLNIGDILVGMKKQKRFVFPQVND 89
 Db 3190 SRRRPLDELPBAQALADAENTTD--AADSAPVLP-----LAGMAAER----- 3234
 QY 90 ATSEKTAKTVPQRITSAILI-----SDPSQOPLAFVNLGFSNTGLQALGITDLDGA- 144
 Db 3235 -----RAWLDLVLAEASIVLGHNGSDPVGPBRAFGELGFDG--LMAVELNRILGAT 3284
 QY 145 --QFPDQFADANLIGDLSQWVAPFTGTTTHGFLISSDDDFLDQTDISTFGSSI 202
 Db 3285 GLSLPATLIFPSPSALAEQLVGLVGAQPATTVVAGADVD-----DEV----- 3330
 QY 203 TQVQALSGSAPFQAGHEHFGFL-----DGISQ-PSVTGME-TTVF-----PGQAVVP 249
 Db 3331 --VVVAMGCRIFGDVCSFEELMOLVSAGRDVSTFPVDRGMDCNTLFPDPDRAGSTYVR 3388
 QY 250 PGIILTRDG-DTG-----TRPSMALDGSFMAFRHFOQKV--DEFNAYTLANAIPANSA 300
 Db 3389 EGARLTGADRFDAGPFGISPREARMDP-----QQRLLEVAWEVEFRAAGIAPLSLR 3440
 QY 301 GNLTOGEAEFLGARMFRKMSGAPIDLAFTADPALGADPQRNNNFYSOTLTDETRCP 360
 Db 3441 GSRT-----GVFAGTN--GQ-DHGAKVAAAPBAAGHLTGN-----ASVLGRLSYT 3485
 QY 361 FGAVRKTNPQDGGP---VDT-----FHMRSSIPYGPERSDAELASGTQAQDRG 409
 Db 3486 FG-----LEGPVAVVDIACSSSLVALHLIACSLRSG--ECDMALGGVTVWSTP 3532
 QY 410 LLEVEY 415
 Db 3533 LAFLEF 3538

RESULT 4

US-08-418-893D-23
 Sequence 23, Application US/08418893D
 Patent No. 5559220
 GENERAL INFORMATION:
 APPLICANT: ROESSLER, PAUL G
 APPLICANT: OHROGGE, JOHN B
 TITLE OF INVENTION: GENE THAT ENCODES ACETYL-COENZYME A
 TITLE OF INVENTION: CARBOXYLASE FROM CYCLOTHELLA CRYPTICA
 NUMBER OF SEQUENCES: 25
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: NATIONAL RENEWABLE ENERGY LABORATORY
 STREET: 1617 Cole Blvd.
 CITY: Golden
 STATE: CO
 COUNTRY: USA
 ZIP: 80401-3393
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/418,893D
 FILING DATE: April 7, 1995
 CLASSIFICATION: 800
 APPLICATION NUMBER: US 08/104,938
 FILING DATE: September 14, 1993
 CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:

NAME: O'CONNOR, EDNA
 REGISTRATION NUMBER: 29,252
 REFERENCE/DOCKET NUMBER: MRI/NREL IR# 92-48CON
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 303-231-1000
 TELEFAX: 303-231-1098
 TELEX:
 INFORMATION FOR SEQ ID NO: 23:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2089 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEICAL: NO
 ANTI-SENSE: NO
 US-08-418-893D-23

Query Match 3.9%; Score 102; DB 1; Length 2089;
 Best Local Similarity 21.0%; Pred. No. 1.6;
 Matches 88; Conservative 53; Mismatches 168; Indels 110; Gaps 21;

QY 9 SVAVLVSSSHVNAKLGARQTRTTLTNFGQAPLPLTLTQHTTESGANDTIL-----PL 64
 Db 1204 NVOIYVSE--KTARRNAQVFLRAISHIPGL-----TTFSGARRALQIGDEL 1251
 QY 65 NNIGDILVGMKKQKRFVFP--QVNDATSPKTA-----LKTVPORI----- 105
 Db 1252 ERAQANSKVSQSSRYLHSLPEOSDAPTEPAIKKEGVIDKLKSLAQRLTLRLRDEI 1311
 QY 106 -TSAAILISD---PSQOPLAFVNLGFSNTGLQALGITDLDGAQFPDQGFADANLIGDD 160
 Db 1312 ETKRVTVQDEDSPRVAVPRLVASSWQGEWLKTSAYIDR-----PD----- 1353
 QY 161 LSOVAPFTGTTTHGVLISDDDFLDQTDISTFGSSITTOVALSGSARFPDQAG- 219
 Db 1354 -----PYTGVRERC-VIGGIDVECELESYSTSTIOTK-RSIAARVSTVAVDYIGL 1405
 QY 220 -----HEHFGFLDISQPSVTGMEITVPFGQAVP--PGIILTRDGDGTRP-----SW 267
 Db 1406 LEVSLGEMDKYLSLSGPDPTIPSNVFEAQELBGPDEGLVTGK-REGITKVGAVAW 1464
 QY 268 ALDGSFMAFRHFOQKVEF---NAYTLANAIPANSAGNLTOGEAEFLGARMFGR----- 319
 Db 1465 VV-----TKTPEYEPGRGVVIVNDVYQGS-GSGVEDEDEVFFASKAKARENKL 1513
 QY 320 -----WKSQAPI---DLAFTADPALG--ADPQRNNNFYSOTLTDETRCPFGAHVRK 367
 Db 1514 PRVVIACNSGARIGLVDDLKPKFQIKFIDEASPSKGFEEYVLDDATYKSLPEGSVNVRK 1572

RESULT 5

US-08-418-893D-24
 Sequence 24, Application US/08418893D
 Patent No. 5559220
 GENERAL INFORMATION:
 APPLICANT: ROESSLER, PAUL G
 APPLICANT: OHROGGE, JOHN B
 TITLE OF INVENTION: GENE THAT ENCODES ACETYL-COENZYME A
 TITLE OF INVENTION: CARBOXYLASE FROM CYCLOTHELLA CRYPTICA
 NUMBER OF SEQUENCES: 25
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: NATIONAL RENEWABLE ENERGY LABORATORY
 STREET: 1617 Cole Blvd.
 CITY: Golden
 STATE: CO
 COUNTRY: USA
 ZIP: 80401-3393
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25

```
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,893D
; FILING DATE: April 7, 1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/104,938
; FILING DATE: September 14, 1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: O'CONNOR, EDNA
; REGISTRATION NUMBER: 29,252
; REFERENCE/DOCKET NUMBER: MRI/NREL IR# 92-48CON
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303-231-1000
; TELEFAX: 303-231-1098
;
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2089 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; US-08-418-893D-24

Query Match 3.9%; Score 102; DB 1; Length 2089;
Best Local Similarity 21.0%; Pred. No. 1.6;
Matches 88; Conservative 53; Mismatches 168; Indels 110; Gaps 21;

QY 9 SVAVLGSSSHVNAKLGARQTRTLLTNFPQAPLTLTQHTESGANDTL-----PL 64
DB 1204 NVOIYVGE---KTARRNAQVFLRAISHTPGI-----TTFGARRALLQGLDEL 1251
QY 65 NNTQGDLLVMKKQKRFVFF--QVNDATSFKTA-----LKTVPORI----- 105
DB 1252 ERAQANSKVSQSSRIYLSLPEQSDATPEETAKEFEGVIDKLKSLAQLRLKLRVDEI 1311
QY 106 -TSAAILISD---PSQQPLAFVNLGFSNTGLQALGTTDLDGDAQFPDGFADAAANLGD 160
DB 1312 ETKVTVTVQEDSGSPVVPVRLVASSMOGELKTSVIDR-----PD----- 1353
QY 161 LSONVAPFTGTHGVFLIGSDODDFLDQFTDDISSTFGSSITQVQALSGSARFPDQAG- 219
DB 1354 -----PVTGVTRERC-VIGEGIDEVCELESYDSTSTIQTK-RSIARRVSGTVAYDYVGL 1405
QY 220 -----HEHFGFLDGISQPSVTGWHTTVFPQAVV--PPGILLTGRDGTGTRP-----SW 267
DB 1406 LEVSLGGEWDKYLSSLSGPDPTTIPSNVFEAQELLEGPDGELVTGK-REIGTNKVGWVAV 1464
QY 268 ALDGSFAPRHFQOKVPEF---NAYTLANAIPANSAGNLTOEGAGFLGARMFGR--- 319
DB 1465 VW-----TWKTEYEGREGVQVVIVNDVTQOS--GSFGVEDEVFVKASKYARENKL 1513
QY 320 -----WKGAPI-----DLAPTADDPALG-ADPQNNNFYSDTLTDETRCPFGAHRK 367
DB 1514 PRVVIACNSGARIGLVDDLKPKFKI1K1DEASPSKGEYLYLDDATYKSLPEGSVNVVK 1572

RESULT 6
5268290-2
; Patent No. 5268290
; APPLICANT: Hasegawa, Mamoru; Sakurada, Kazuhiro
; TITLE OF INVENTION: PROCESS FOR PRODUCING NEURAMINIDASE
; NUMBER OF SEQUENCES: 2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/384,709
; FILING DATE: 25-JUL-1989
; SEQ ID NO: 2
; LENGTH: 626
5268290-2

Query Match 3.9%; Score 101.5; DB 6; Length 626;
Best Local Similarity 20.2%; Pred. No. 0.27;
Matches 102; Conservative 45; Mismatches 145; Indels 213; Gaps 28;

QY 116 SQOPLAFVN--LQFSNTGLQALGITDLDGDAQFPDQCFADAAANLGDLDLQSWVAPFTGTTI 173
DB 15 TEQDLA-VNREGFPNRYRIPALTIVTPD-----GDLASYDGRPTGIDA 56
QY 174 HGVFLIGSDODDFLDQFTDDISSTFGSSITQVQALSGSARFPDQAGHEHFGFLDGISQPS 233
DB 57 PG-----PNSILQRRSTGGRTWGEQ--QVVSAGOTTAP-----IKGFSDES 96
QY 234 -VTGWET-TVF-----PGQAVVPPGI-----ILTGRDG----- 259
DB 97 YLVDRGTGTFINPHVYSQRQGFAGSRPGTDPADPNVLHANVATSTDGLTWSHRTITADI 156
QY 260 --DTGTRPSWALDGSFMAFRHFQOKVPEFNAYTLANAIPANSAGNLTOEGAGFLGARMF 317
DB 157 TPDGWRSRFAASGEGIQRYGPHAGRLTQQYTIINAAGAFQAVSVYSDD-----H 207
QY 318 GR-WKSGAPIDLAPTADDPALGADPQNNNFYSD-----TLTDETRCPFGAHRKTNPR 371
DB 208 GRTWRAG-----BAVGVMDEKTVLSDGRVLLNSRDSARSGY---RKVAVS 252
QY 372 QDLGGPVDTFHAMRSSIPYGPETSDAEL-----ASGVTAQDRGLLF----- 412
DB 253 TD-GG-----HS-----YGPVTIDRDLDPDTNNASIRAPDPAPAGSARAKVLLFSNA 300
QY 413 -----VEYQSI-----GNGFRFQQIN 429
DB 301 SQTSRQGTIRMSCDGQGTWPVSKVPQGSMSYSTUTALPDGTYGLLYEPGTQIRYANFN 360
QY 430 --WANNANFFFSKP---ITPGIEPII-----GOTTPR-----TVGGLDPL 464
DB 361 LAWLGGICAPFTIPDVALEPGQQTVPVAVTQSGIAVPKPSLQLDASPDWQVQGSVEPL 420
QY 465 NQNE-----TFTVPLFVIPKGEY 483
DB 421 MPGRQAKQVTTITVPAGTTP--GRY 443

RESULT 7
US-07-743-357-10
; Sequence 10, Application US/07743357
; Patent No. 5858646
; GENERAL INFORMATION:
; APPLICANT: Kang, Yong C.
; TITLE OF INVENTION: Polypeptide having immunological
; activity for use as diagnostic reagent and/or vaccine
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KIRBY EADES GALE BAKER
; STREET: Box 3432, Station D
; CITY: Ottawa
; STATE: Ontario
; COUNTRY: Canada
; ZIP: K1M 1H8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/743,357
; FILING DATE: 21-AUG-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA90/00062
; FILING DATE: 23-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gale, Edwin J.
; REGISTRATION NUMBER: 28,584
```

REFERENCE/DOCKET NUMBER: 30924-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (613) 237-6900
TELEFAX: (613) 237-0045
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1003 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHEICAL: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
ORGANISM: Human immunodeficiency virus type 1
STRAIN: ELI
US-07-743-357-10

Query Match 3.8%; Score 100; DB 2; Length 1003;
Best Local Similarity 22.1%; Pred. No. 0.8;
Matches 76; Conservative 41; Mismatches 117; Indels 110; Gaps 18;

QY 85 FOVNDATSFKAL-KTVVPORTSAALIDPSQO-PLAFVNLGFSNTGLQALGITDD 140
DB 594 FVYGAANRETKLKAQVYTRGRQKVPLDTTNQKTELQALIDPSGLVAVITDS 653
QY 141 -LGDAQF-PDQGFADAN-LGDDLSQ-----WVAFTGTHGVFLIGSDDDFLD 188
DB 654 QYALGITIQAQPKSESELVNOIIEOLIKKEYLAVPAHKG-----IGNEGVDLK 705
QY 189 QFTDISSTFGSSITQVALSGSARPFDOAGHEH-----FGFLDGISQPSVT 235
DB 706 -----VSGIRKVLFLDG-----IDKAQEHKHYNNRRAMASDNLPPVAKETVA 752
QY 236 GWETTFFGQAV-----VPRGILTRDGDGTGRPSMALDGSFMAFRFPQKVFENAYT 290
DB 753 SCDCQKLGEMHGOVDCSPGI-----WOLDCTHLEKXILVAVHVASGYI 798
QY 291 LANIPANSANLTOQEGAEFLGAMFGRWKSGAPIDLAPTADDALADPPQRNNFDYS 350
DB 799 EAELVPAE-----TGQETAFVL-LTLAQRW-----PVKVVHT-----DNGSNF--- 835
QY 351 DTLDETRCFPGAHVKTNPRLDGGPVDTFHAMRSSIPYGPET 394
DB 836 -----TSAVRAKACWAGIKQEPG-----IPYNQS 861

RESULT 8
US-08-415-788-7
Sequence 7, Application US/08415788
Patent No. 5834591
GENERAL INFORMATION:
APPLICANT: NORMARK, STAFAN
APPLICANT: JONSSON, ANN-BETH
TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES USEFUL
TITLE OF INVENTION: FOR THE DIAGNOSIS AND TREATMENT OF PATHOGENIC NEISSERIA
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08415.788
FILING DATE:
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/829,465
FILING DATE: 31-JAN-1992
ATTORNEY/AGENT INFORMATION:
NAME: SCHRENNING, LYNN E.
REGISTRATION NUMBER: 37,233
REFERENCE/DOCKET NUMBER: 29500-20046.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 998 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-415-788-7

Query Match 3.8%; Score 98.5; DB 2; Length 998;
Best Local Similarity 20.5%; Pred. No. 1.1;
Matches 120; Conservative 55; Mismatches 200; Indels 209; Gaps 30;

QY 24 KLGARQRTTTLTNFPG-----QAPLETLTQHTTESG-----ANDTLPLNNI 67
DB 279 KLGLPQVK-AGRCITNKPRVNNNTKAPSPALFAPALMFGQDGAEMYSASVSTYPSDSS 337
QY 68 QGDLVLGKKQKE-----RFVFOVNDATSFKTLKTYVPO-----RTSAAILI-- 112
DB 338 SRIFQELKTYTEBQKPKQRYSLKSLNDG-----BIKSRQSPFNGRQTIIRLDGVLHLKL 392
QY 113 SDPSQQLPFAVNLGFSNTGL-QALGITDDLGDAQFPDQGFADANLGDLSQWVAFPTGT 171
DB 393 NSKDEVAAPFVNLGNNTGKNDTGTGYKE-----ANNLIDAEKKVLLPW 438
QY 172 TIHGVFLIGSDQDDFLDQFTDISSTFGSSITQVALSGSARPFDOAGHEHFGFLDGISQ 231
DB 439 TVRG-----PDNNKFKSINQKPEKYSQRYRINDNNGNRDLG--DIYNS 480
QY 232 P-SYTGWETT-----FG-----QAVVPQIT 253
DB 481 PIVAGGYLATANDGWVHIFKNGGSDERSYNLKL5Y1PCTMRKDIQSGESTLAKELR 540
QY 254 LTRDGDGTGRPSMALDGSFMAFR--HFQKVPEF-----NAVYLANAIPANSAGN 302
DB 541 APAEKGYVGR--YGVDSGFVLRQVELSGQKHVFMFGMCGGCAVAL-----D 588
QY 303 LTOQGAFFLGARME-----GRWKSAPIDLAPTADDPALG----- 338
DB 589 LSKINGVPAAPLFDVADGDNNGNKRKVELGYTVGTPQIKIRNGKYAFLASGYAAK 648
QY 339 -ADFORNN--FDYSTTLDETTRCFPGAHVKTNPRLDGGPVDTFHAMRSSIPYGPET 393
DB 649 KIDSTNTKALYVVDLKTTL-----GTPIAKIEVXDGGK-----LSS-----PT 688
QY 394 TSDPELASGVT--AODGGLFVEYQSIIGGFRQOIN-----WANNANPFSPKPIPG 445
DB 689 LVYDKLDGCTVDIAVAGDRG-----GMYRDRDLSNSSSKWSAKAVIFEGDKPITSA 738
QY 446 IEPILGQTTPTTV-----GGI-----DPLANONETFTVPLFVPIKG 480
DB 739 --PAVSRLADKRVIYFGTSGDLTDEDDVLTGEOYIYGFIDDDKG 780

RESULT 9
US-07-743-357-7
Sequence 7, Application US/07743357
Patent No. 5856646
GENERAL INFORMATION:
APPLICANT: Kang, Yong C.
TITLE OF INVENTION: Polypeptide having immunological
TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine
NUMBER OF SEQUENCES: 22

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: KIRBY EADES GALE BAKER
STREET: Box 3432, Station D
CITY: Ottawa
STATE: Ontario
COUNTRY: Canada
ZIP: K1M 1H8

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/743,357
FILING DATE: 21-AUG-1991
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/CA90/00062
FILING DATE: 23-FEB-1990
ATTORNEY/AGENT INFORMATION:
NAME: Gale, Edwin J.
REGISTRATION NUMBER: 28,584
REFERENCE/DOCKET NUMBER: 30924-2
TELEPHONE: (613) 237-6900
TELEFAX: (613) 237-0045
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1004 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
ORGANISM: Human immunodeficiency virus type 1
STRAIN: SF2
US-07-743-357-7

```

```

Query Match      3.8%; Score 98.5; DB 2; Length 1004;
Best Local Similarity 22.0%; Pred. No. 1.1; Indels 113; Gaps 20;
Matches 78; Conservative 49; Mismatches 114;

QY 78 QKRFV---FFQVNDATSPKTL--KTYVPQRITSAAILISDPSQ--PLAFVNLGFSNT 130
DB 585 EKEPIVGAETFFYDGAANRETKLGAGYVTDRCQKVSIADTTNOKTELQAHALQDS 644
QY 131 GLQALGITDD---LGAQF-PGQFAD-AANLGGDLSQ-----WVAPPTGTHGVFL 178
DB 645 GLEVNIVTDSQVAGIQAQPKDSELSVQIIIEQLIKKEKVKYLAWVPAHKG-----I 697
QY 179 IGSDDDFLDQFDTDLSSTFGSSITQVQALSGSARPPDOAGHEH-----FGF 225
DB 698 GGNEQVDKL-----VSAGIRKVFUNG-----IDKAEHEKHYSHWRAMASDFNL 743
QY 226 LQISQPSVTGWETTVPQAV-----VPPGIILTRGDGTGRPSWALDGSFMAFRHFQ 280
DB 744 PPVAKIEIVASCDKCKLGEAMHGVDSCPGI-----WQDCTHLECKIIL 789
QY 281 QKVPEFNATLANAIPANSAGNIQTQEGAEFLGARFGRKWSGAPIDLAPTADDPALGAD 340
DB 790 VAVHVASGVIEAEVPAE-----TGOETAYFL-LKLAGRW-----PVKTIHT----- 830
QY 341 PQNNNFYSDTLTDTETRCFPGAHVRKTNPRDLGGPVDTFHAWRSSIPVGPET 394
DB 831 ---DNGSNFTST-TVKAACW-----AGIKQBEFG-----IPYNPOS 862

```

```

RESULT 10
US-08-415-788-3
; Sequence 3, Application US/08415788
; Patent No. 5834591

```

```

GENERAL INFORMATION:
APPLICANT: NORMARK, STAFFAN
APPLICANT: JONSSON, ANN-BETH
TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES USEFUL
TITLE OF INVENTION: FOR THE DIAGNOSIS AND TREATMENT OF PATHOGENIC NEISSERIA
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/415,788
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/829,465
FILING DATE: 31-JAN-1992
ATTORNEY/AGENT INFORMATION:
NAME: SCHWENNING, LYNN E.
REGISTRATION NUMBER: 37,233
REFERENCE/DOCKET NUMBER: 29500-20046.20
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1029 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-415-788-3

```

```

Query Match      3.8%; Score 98.5; DB 2; Length 1029;
Best Local Similarity 20.5%; Pred. No. 1.2; Indels 209; Gaps 30;
Matches 120; Conservative 55; Mismatches 200;

QY 24 KLGARQTRRTPLLTNPPG-----QAPLPLTLQHTTESG-----ANDTILPLNNI 67
DB 310 KLGLPQVK-AGRCTNKPNNNTKAPSPALAPALWFGPGQDGKAEYASVSTYDSSS 368
QY 68 QGDILVGMKKQKE-----RFVFFQVNDATSFKTALKTYVPQ-----RITSAAILI-- 112
DB 369 SRIFLQELKTQTEPGKPGRYSLKSLNDG-----EIKSRQPSFNGRQTIIRLDDGVHLKL 423
QY 113 SPFSQOPLAFVNLGFSNTGL-QALGITDLDLGDQAPDQGFADAAANLGGDLSQWVAFPTGT 171
DB 424 NGSKDVAFAFVNLGNNTGKNTDFTGIVKE-----ANVNLDADEWKKVLLPW 469
QY 172 TIHGVELIGSDQDDDFLDQFDTDLSSTFGSSITQVQALSGSARPPDOAGHEHFGFLDGIQ 231
DB 470 TVRG-----PNDNKFKSINOKPEKYSORYIRDNNGNRDLG--DIVNS 511
QY 232 P--SVTGWETT-----FPG-----QAVVPPGI 253
DB 512 PIVAVGGYLATAANDGMVHIFKKNGGSDERSYNLKLSTPGTMRKDIQSQESTLAKELR 571
QY 254 LTGRDQDTGRPSWALDGSFMAFR---HFQOKVPEP-----NAYTLANALPANSAGN 302
DB 572 AFAEKYGVQDVR--YGVDDGGFVLRQVELSQCKHVFMFANGFGGGRGAYAL-----D 619
QY 303 LTOQEGAEFLGARMF-----GRWKS GAPIDLAPTADDPALG----- 338
DB 620 LSKINGNYFAAAPLFDVKDGDNDNGNRKVRKVELGYTGTGTQIGKIRNGKYAAFLASGYAAK 679

```



```
QY 339 -ADPQRNN-----FDYSDLTLDTRCPFGAHVRKTNPRDLGCPVDTFHAMRSSIPYGPPE 393
DB 680 KIDSTKRLALYVLDLDTL-----GTPIAIEVDGKGG-----LSS-----PT 719
QY 394 TSDAELASGVY---ADRGLLFVEYQSLIGCFEPQOIN-----MANNANPEFKPIITPG 445
DB 720 LVYDLDLGGYDIAYAGDGG-----GMMYTFDLSNSSSKMSAKVITEGDKPIITSA 769
QY 446 IEPILGQTTPTV-----GGL---DPLNQNFTFVPLFVPIPKG 480
DB 770 --PAVSRLLADKRVYIFGTGSDLTEDDVLTNGEQYIYGFDDDKG 811

RESULT 11
US-07-743-357-22
; Sequence 22, Application US/07743357
; Patent No. 5858646
; GENERAL INFORMATION:
; APPLICANT: Kang, Yong C.
; TITLE OF INVENTION: Polypeptide having immunological
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KIRBY EADES GALE BAKER
; STREET: Box 3432, Station D
; CITY: Ottawa
; STATE: Ontario
; COUNTRY: Canada
; ZIP: K1M 1H8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/743.357
; FILING DATE: 21-AUG-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA90/00062
; FILING DATE: 23-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gale, Edwin J.
; REGISTRATION NUMBER: 28,584
; REFERENCE/DOCKET NUMBER: 30924-2
; TELEPHONE: (613) 237-6900
; TELEFAX: (613) 237-6900
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Human immunodeficiency virus type 1
; STRAIN: HXB2
; US-07-743-357-22

Query Match 3.7%; Score 97.5; DB 2; Length 913;
Best Local Similarity 22.1%; Pred. No. 1.2; Indels 121; Gaps 19;
Matches 79; Conservative 42; Mismatches 116; Indels 121; Gaps 19;
```

```
QY 179 IGSDDDFLDQFTDIDISTFGSSITVOVALSGSARPPDOAGHEH-----FGF 225
DB 607 GGNQVDKX-----VSAGIRKVLFLDG-----IDKQDDEHEKXHSNWRAMASPNL 652
QY 226 LDGISOPSVTCMETTVPFGQAV-----VPPGIIITGADGDTGTRPSMALDGSFMAFHHFO 280
DB 653 PPVAKETIVASCDCQKCGKGEAMHGQVDCSPGI-----WQLDCTHLEGKYL 698
QY 281 QKVPENAYTLTANAI PANSAGNLTOGGAFLGARMFGRWKSAPIDLAFTADDPALGAD 340
DB 699 VAVVAGSYIEAEVILPAE-----TGQETAYFL-LKLAGRW----- 732
QY 341 FORNNPFYSPTLTDETRCPFGAHVRK-----TNPRDLGCPVDTFHAMRSSIPYGPET 394
DB 733 PVKTIHNDGNSNFT-----GATVRAACWMAIGIKQFEG-----IPYNPOS 771

RESULT 12
US-07-743-357-1
; Sequence 1, Application US/07743357
; Patent No. 5858646
; GENERAL INFORMATION:
; APPLICANT: Kang, Yong C.
; TITLE OF INVENTION: Polypeptide having immunological
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KIRBY EADES GALE BAKER
; STREET: Box 3432, Station D
; CITY: Ottawa
; STATE: Ontario
; COUNTRY: Canada
; ZIP: K1M 1H8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/743.357
; FILING DATE: 21-AUG-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA90/00062
; FILING DATE: 23-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gale, Edwin J.
; REGISTRATION NUMBER: 28,584
; REFERENCE/DOCKET NUMBER: 30924-2
; TELEPHONE: (613) 237-6900
; TELEFAX: (613) 237-6900
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1005 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Human immunodeficiency virus type 1
; STRAIN: HXB2
; US-07-743-357-1

Query Match 3.7%; Score 97.5; DB 2; Length 1005;
Best Local Similarity 22.1%; Pred. No. 1.4; Indels 116; Gaps 19;
Matches 79; Conservative 42; Mismatches 116; Indels 121; Gaps 19;
```

Db 586 EKEPIVCAETFFYVDGAANRETKLGAGYVTRGRQKVVTLTDTTNKTELQAIYLAQDS 645
QY 131 GLQALGITDD---LGAQF-PDQGFADAAN-LGDDLSQ-----WVAPTGTTHGVFL 178
Db 646 GLEVNIVTDSQVALGIITQAPDQSESELVNOIIEQLIKKEKVLAWVPAHKG-----I 698
QY 179 IGSDDDFLDQFTDDISSTFGSSITOVQALSGSARPPDOAGHEH-----FGF 225
Db 699 GGNQEQVDKL-----VSAGIRKVLFDG-----IDKAQDEHEKXHSNWRAMASDFNL 744
QY 226 LOGISOPSVTGWETVFPQAV-----VPPGIILTGRDGTGTRPSWALDGSFMAFRHFQ 280
Db 745 PPVVAKEIVASCDKQLKGEAMHGQVDCSPGI-----WQLDCTHLEKGVIL 790
QY 281 QKVPEFNAYTLANAIPANSAGNLTOQEGAEFLGARFGRWKSGAPIDLAPTADDPALGAD 340
Db 791 VAVHVASGYIEAEVPAE-----TGQETAYFL-LKLAGRW----- 824
QY 341 PQRNNFDYSDTLTDETRCPFGAHRK-----TNPRQDLGGPDTFFHAMRSSIPYGPET 394
Db 825 PVKTIHDTNGSNFT-----GATVRAACWAGIKOEF-----IPYNPOS 863

RESULT 13
US-07-743-357-9
; Sequence 9, Application US/07743357
; Patent No. 5858646
; GENERAL INFORMATION:
; APPLICANT: Kang, Yong C.
; TITLE OF INVENTION: Polypeptide having immunological
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KIRBY BADES GALE BAKER
; STREET: Box 3432, Station D
; CITY: Ottawa
; STATE: Ontario
; COUNTRY: Canada
; ZIP: K1M 1H8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/07743,357
; FILING DATE: 21-AUG-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA90/00062
; FILING DATE: 23-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gale, Edwin J.
; REGISTRATION NUMBER: 28,584
; REFERENCE/DOCKET NUMBER: 30924-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (613) 237-6900
; TELEFAX: (613) 237-0045
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1003 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: Internal
; ORIGINAL SOURCE:
; ORGANISM: Human immunodeficiency virus type 1
; STRAIN: MAL
US-07-743-357-9
Query Match 3.7%; Score 97; DB 2; Length 1003;

Best Local Similarity 21.8%; Pred. No. 1.6;
Matches 75; Conservative 43; Mismatches 116; Indels 110; Gaps 18;
QY 85 FQVNDATSKT--ALKTYVPQRITSAAIILSPSOQ--PLAFVNLGFSNTGLQALGITDD 140
Db 594 FYVDGAANRETKKGAGYVTRGRQKVVTLTDTTNKTELQAIHLALQDSGLEVNIVTDS 653
QY 141 ---LGAQF-PDQGFADAAN-LGDDLSQ-----WVAPTGTTHGVFLIGSDQDDFLD 188
Db 654 QVALGIITQAPDQSESELVNOIIEQLIKKVLWSVPAHKG-----IGNEQVDKL- 705
QY 189 QFTDDISSTFGSSITOVQALSGSARPPDOAGHEH-----FGFLDGIOSPSVT 235
Db 706 -----VSSGIRKVLFDG-----IDKAQDEHEKXHSNWRAMASDFNLPPVAKI 752
QY 236 GWETVFPQAV-----VPPGIILTGRDGTGTRPSWALDGSFMAFRHFQKVPFNAYT 290
Db 753 SCDKQLKGEAMHGQVDCSPGI-----WQLDCTHLEKGIILVAVHVASGYI 798
QY 291 LANAIPANSAGNLTOQEGAEFLGARFGRWKSGAPIDLAPTADDPALGADPQRNNFDYS 350
Db 799 EAEVPAE-----TGQETAYFI-LKLAGRW-----PVKVVHT-----DNGSNP--- 835
QY 351 DTLTDETRCPFGAHRKVTNPRQDLGGPDTFFHAMRSSIPYGPET 394
Db 836 -----TSAAVRAACWAGIKOEF-----IPYNPOS 861
RESULT 14
US-08-485-355B-50
; Sequence 50, Application US/08485355B
; Patent No. 6177075
; GENERAL INFORMATION:
; APPLICANT: Christian, P. D., Gordon, K. H.J., Hanzlik, T. N.
; TITLE OF INVENTION: Insect Viruses and Their Uses in
; Protecting Plants
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,355B
; FILING DATE: 07-Jun-1995
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/440,522
; FILING DATE: 12-MAY-1995
; APPLICATION NUMBER: US 08/089,372
; FILING DATE: 08-JUL-1993
; APPLICATION NUMBER: AU PL4081/92
; FILING DATE: 14-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard P.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-58631-2/RFT/DSS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 647 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-08-485-3558-50

Query Match 3.7%; Score 96; DB 4; Length 647;

Best Local Similarity 21.0%; Pred. No. 0.98; Mismatches 198; Indels 126; Gaps 21;
Matches 99; Conservative 49;

58 NDTLPLNNIIGDILVGMKKQKRRFVFOVNDATSPFTALKTYPORITSAAIISPSQ 117
184 NDLEIWMNNLADMRV---VDSEQMINF-TNDTT-----YYRIKIVLRTYVPPTE 232
118 QPLAFVNLGFSNTGLQALGITDDIGDAQFPDGFADANLGDLSQWVAPFTGTHGVF 177
233 -----GLVRTVSDYRLTYKAITCEANMPTLVQ-----GFWIGQY 268
178 LIGSDQDDFLDQFTDDISSTFG--SSITQVQALSGSARFPDQAGHFGFLDGISQPSVTG 236
269 ALTPPTS---LPQY--DVSEAYALHTLTFARPSSAALAFAVWAGLPQ---GTAAPACTPA 319
237 WET-----TVFPGA---VVPFGIILTGRDGTGTRPSWALDGSFMAFRH-- 278
320 WEQASSGGYLTWRNNGTTFPAGSVSYLPEGFALERIDPRDGSMTDPAASGDTVTFQVA 379
279 -----FOQVPEFNAAYTLANAIPANSAGNLTOQEGAEFLGARMFGRW 320
380 VDEVVVTNPNAGGGSAPLTFVRVPPSNAYT--NTVFRNTL--LETRSSRLLELPM---- 431
321 KSGAPILAPLADDPALGADPQRNNNFDYSDTLTDETRCPFGAHRKTNPRQDGGVDT 380
432 ---PPADFGQTV-----ANNPKIEOSLKETLGCYLVHSKXRNVPFOL--TPASS 476
381 FHAMRSSIPYGPETSDAELASGVTADRGLLFVEYOSIIGNFR-----FQOI 428
477 FGAVSPNPGYERTRDLPDTGIRDS-----FDQNMSTAANHFRLSHSCSIYTKTYQW 531
429 NWANNANPFPSKPIIPGI---EPI--IGQTPRTVGGIDPLNQNERTVPLF 475
532 EGVTVNVTTPFGQFAHAGLKNKEILCLADDLATRLTGVPATDPAVAASAF 583

RESULT 15
US-09-194-613-3
Sequence 3, Application US/09194613
Patent No. 6251654
GENERAL INFORMATION:
APPLICANT: GORDON, Karl H.
APPLICANT: HANZLIK, Terry N.
TITLE OF INVENTION: MODIFIED SMALL RNA VIRUSES
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDermott, Will & Emery
STREET: 600 13th Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3096
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/194,613
FILING DATE: 30-NOV-1998
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Bucca Ph.D., Daniel
REGISTRATION NUMBER: 42,368
REFERENCE/DOCKET NUMBER: 50179-061
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-756-8000
TELEFAX: 202-756-8087

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 647 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-194-613-3

Query Match 3.7%; Score 96; DB 4; Length 647;

Best Local Similarity 21.0%; Pred. No. 0.98; Mismatches 198; Indels 126; Gaps 21;
Matches 99; Conservative 49;

58 NDTLPLNNIIGDILVGMKKQKRRFVFOVNDATSPFTALKTYPORITSAAIISPSQ 117
184 NDLEIWMNNLADMRV---VDSEQMINF-TNDTT-----YYRIKIVLRTYVPPTE 232
118 QPLAFVNLGFSNTGLQALGITDDIGDAQFPDGFADANLGDLSQWVAPFTGTHGVF 177
233 -----GLVRTVSDYRLTYKAITCEANMPTLVQ-----GFWIGQY 268
178 LIGSDQDDFLDQFTDDISSTFG--SSITQVQALSGSARFPDQAGHFGFLDGISQPSVTG 236
269 ALTPPTS---LPQY--DVSEAYALHTLTFARPSSAALAFAVWAGLPQ---GTAAPACTPA 319
237 WET-----TVFPGA---VVPFGIILTGRDGTGTRPSWALDGSFMAFRH-- 278
320 WEQASSGGYLTWRNNGTTFPAGSVSYLPEGFALERIDPRDGSMTDPAASGDTVTFQVA 379
279 -----FOQVPEFNAAYTLANAIPANSAGNLTOQEGAEFLGARMFGRW 320
380 VDEVVVTNPNAGGGSAPLTFVRVPPSNAYT--NTVFRNTL--LETRSSRLLELPM---- 431
321 KSGAPILAPLADDPALGADPQRNNNFDYSDTLTDETRCPFGAHRKTNPRQDGGVDT 380
432 ---PPADFGQTV-----ANNPKIEOSLKETLGCYLVHSKXRNVPFOL--TPASS 476
381 FHAMRSSIPYGPETSDAELASGVTADRGLLFVEYOSIIGNFR-----FQOI 428
477 FGAVSPNPGYERTRDLPDTGIRDS-----FDQNMSTAANHFRLSHSCSIYTKTYQW 531
429 NWANNANPFPSKPIIPGI---EPI--IGQTPRTVGGIDPLNQNERTVPLF 475
532 EGVTVNVTTPFGQFAHAGLKNKEILCLADDLATRLTGVPATDPAVAASAF 583

Search completed: November 27, 2002, 13:46:26
Job time: 27 secs

•
•
•
•

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 27, 2002, 13:46:05 ; Search time 13 Seconds
(without alignments)
610.019 Million cell updates/sec

Title: US-09-926-084-7
Perfect score: 2601
Sequence: 1 MSLSLFVSVAVLVGSSSHV.....KGGYFPLPSALTATIAA 498

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 102317 seqs, 15924203 residues
Total number of hits satisfying chosen parameters: 102317

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
- 2: /cgn2_6/ptodata/2/pubppaa/PCR_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*
- 5: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
- 6: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubppaa/PCRUS_PUBCOMB.pep:*
- 8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep:*
- 10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep:*
- 12: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
- 13: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*
- 14: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	137.5	5.3	439	10 US-09-815-242-11712	Sequence 11712, A
2	130.5	5.0	423	10 US-09-815-242-10121	Sequence 10121, A
3	118	4.5	2383	10 US-09-812-020-303	Sequence 302, App
4	105	4.0	753	10 US-09-815-242-10181	Sequence 10181, A
5	101	3.9	398	10 US-09-815-242-5469	Sequence 5469, Ap
6	101	3.9	409	10 US-09-815-242-12418	Sequence 12418, A
7	96	3.7	1349	10 US-09-815-242-5898	Sequence 5898, Ap
8	96	3.7	1349	10 US-09-815-242-13137	Sequence 13137, A
9	95	3.7	991	10 US-09-815-242-5803	Sequence 5803, Ap
10	92.5	3.6	978	12 US-10-007-693-65	Sequence 65, Appl
11	92	3.5	349	10 US-09-919-935-2	Sequence 2, Appl
12	91	3.5	707	10 US-09-764-870-313	Sequence 313, App
13	90	3.5	637	10 US-09-852-053-3	Sequence 3, Appl
14	90	3.5	931	10 US-09-779-081-2	Sequence 2, Appl
15	90	3.5	1609	12 US-10-007-693-75	Sequence 75, Appl
16	89.5	3.4	646	10 US-09-777-710A-1	Sequence 1, Appl
17	89.5	3.4	670	10 US-09-777-710A-15	Sequence 15, Appl
18	89.5	3.4	3472	9 US-10-027-806-4	Sequence 4, Appl
19	88	3.4	376	10 US-09-815-242-10160	Sequence 10160, A

20	88	3.4	870	10 US-09-815-242-5493	Sequence 5493, Ap
21	88	3.4	870	10 US-09-815-242-12637	Sequence 12637, A
22	87.5	3.4	628	10 US-09-815-628-2	Sequence 2, Appl
23	87.5	3.4	5215	9 US-09-860-846-2	Sequence 2, Appl
24	87.5	3.4	5215	10 US-09-861-289-2	Sequence 12, Appl
25	87	3.3	529	10 US-09-731-393-12	Sequence 57, Appl
26	87	3.3	1047	10 US-09-866-562-57	Sequence 13249, A
27	86	3.3	334	10 US-09-815-242-13249	Sequence 4, Appl
28	86	3.3	572	9 US-09-942-185-4	Sequence 17, Appl
29	86	3.3	1477	9 US-10-092-880-4	Sequence 3, Appl
30	85	3.3	516	10 US-09-731-393-17	Sequence 3, Appl
31	84	3.2	333	10 US-09-897-898-3	Sequence 4, Appl
32	84	3.2	333	10 US-09-897-898-4	Sequence 14, Appl
33	84	3.2	338	10 US-09-897-898-14	Sequence 19, Appl
34	84	3.2	388	10 US-09-897-898-17	Sequence 4, Appl
35	84	3.2	451	10 US-09-897-898-19	Sequence 104, App
36	84	3.2	568	12 US-10-080-223-4	Sequence 11, Appl
37	84	3.2	1337	10 US-09-801-368-104	Sequence 446, App
38	83.5	3.2	716	9 US-10-036-492-11	Sequence 10, Appl
39	83.5	3.2	722	10 US-09-853-533A-10	Sequence 5, Appl
40	83	3.2	660	10 US-09-841-133-446	Sequence 12996, A
41	83	3.2	887	10 US-09-815-242-13179	Sequence 5, Appl
42	83	3.2	996	8 US-08-910-386A-5	Sequence 48, Appl
43	83	3.2	6281	10 US-09-815-242-12996	
44	82.5	3.2	443	12 US-10-083-452-5	
45	82.5	3.2	611	10 US-09-829-549A-48	

ALIGNMENTS

RESULT 1
US-09-815-242-11712
Sequence 11712, Application US/09815242
Patient No. US20020601569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.01A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11712
LENGTH: 439
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-815-242-11712
Query Match 5.3%; Score 137.5; DB 10; Length 439;
Best Local Similarity 21.2%; Pred. No. 7.9e-05;


```

QY 106 TSAAILSDSQOPLAFVNLGFS-----NTGL---QALG----- 136
Db 1086 TTLTAIVKDSNHFVAGIYVNFMTPODVAAANFTLLENGGAIITQANGAEHAVTLKGGKAGTH 1145
QY 137 -ITDDLDGAOPPGQ-----FADAAINGDDLSQWVAEFTTTHGVFLIGSDODDFLDQFT 191
Db 1146 TVTATLTGNNTSDSQPVTFFADAKASQVVLQISKEBITGNGVDSATLTATVYKDP-DNEV 1204
QY 192 DDISTFTGSSITTOVALISGSGARPPDOAGHEHGFGLDGISQPSVTGWETTPPGQAVVPPG 251
Db 1205 NNLPVTFSSASSGILLTPGVSNNTNE-----SGIAQATILAG-----VAFGEKTVTAS 1250
QY 252 IILTRGRODGTGRPSMVLDSFMAFHFQOKVPEFPAATYLTAAIAPANSAGNLTQEGAEF 311
Db 1251 LANNAS-----DKRTVHFIQDTAAKI----- 1273
QY 312 LGAMFGRMKSGAPIDLPAPDADPALGADPQRNNNFEDYSDTLTDETRCP----- 361
Db 1274 -----IELAPPPDSTIAGT-PQNSSGSVITATVYDNNNGFPVKGYTNFTSN 1318
QY 362 GAHVKTNPQR---DLGSGVDTFPHMRSSIPYG--PETSDELASG----- 402
Db 1319 AATAEMTNGGQAVTNEGKATVYTNTRSSIESGARPDIVEASLENGSSTLSTISINVAD 1378
QY 403 -----VTAQGRGLLPVEYQSIITNGRFRFOIMANNANPPSPKPIPTBIE 447
Db 1379 ASTAHLTLQALPEYVASGETTSLTYIEVKDNYNGVPODEVLTSVP-----SEGTPPSNN 1434
QY 448 PIIQOTTPRTVGGDPLNQNERTFVPLFVLPKGEY-----FLPSPISA 491
Db 1435 AI--YTT-----NHGDNFYAS-FTATKAGVYQLTATLLENGDSMOQVTVYVFNVAN 1481
QY 492 LTATIAA 498
Db 1482 AEITLAA 1488

RESULT 4
US-09-815-242-10181
; Sequence 10181, Application US/09815242
; Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
FILE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/0191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: PasteSeq for Windows Version 4.0
SEQ ID NO 10181
;
; LENGTH: 7531
;

```

```

TYPE: BRT
ORGANISM: Escherichia coli
US-09-815-242-10181

Query Match
Best Local Similarity 19.9%, Pred. No. 0.16
Matches 123; Conservative 55; Mismatches 205; Indels 234; Gaps 31

4.0%: Score 105; DB 10; Length 753;

OY 40 PGQADPLTLTQHTTESG--ANDTILPLNNIGDILVGKKOKERFVFO-----VNDAT 91
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 38 PAABPTPGADPTAGSIKAPDT-----BNEKLNSLEDVRKSGSEYALATNQGVRADDO 93
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 92 SFK-----TALKTY-----VPRQITSA-----AII 112
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 94 SLRAGSRGPTLLEDFIREKITHFDHERI PEKIVHARGSAHGYOPYKSLSDITKADFL 153
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 113 SDPSGQPLAFVNL-----GRSNTGLQALGITDGLG-----DA-Q 145
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 154 SDPKKITVFVRFSTVQAGAGSADTVDRIDRGATKFTYEEGIFDVGNNTPIFFIODAK 213
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 146 PEDGQFADANLGDLSQWVAFTGTTHGVFLIGSDDDFLDQETDIDISSTFGSSITGV 205
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 214 FPD--FVHAVK---PEHWAIP-QGQSAH-----DTFMDVYSLQETLHNW 254
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 206 QALSSARPFDOAGHNGF-----LDGISQBSVTGHTTVFQQAIVVPGIILTGRD 258
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 255 WAMSDRGIPRSYRTMEGFIGITFLRLNAEGKATFVRFHKKPLAGKASLVWDEAKLTGRD 314
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 259 GDTGRPSW-ALD-GSFMAFRHFOQKVPEFNAYTL-----ANAI PANSAGNLT 304
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 315 PDFHRELMEALREADDFEYELGFOILFEDEDFKDFOLDLPTKILPELVVYORGVKNV 374
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 305 QOEGAEFLGARMFGKWSGA--PIDLAPTAD--DPALGADPQRNNNPDYSDTLTDETRC 359
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 375 LNRNPD-----NFAFNEQAQAFHGHIVGLDPTNPL-----QGRLFSTDTQISR 422
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 360 PFGAIVRKTNPQDLCGPDVTTHAM--RSSLPY-----GPEISDA----- 397
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 423 -----LGGP--NFHEIPINRPTCYHNHFORGHNRMGIDINPNANPEPNSIN 466
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 398 -----ELASGVTDNRGLLFEVYOS-----IIGNGFRQOIMWANN 433
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 467 DNMPEPTPPGPRGGESEYQERVEGKVRERSPSGEYSHPRLFWLSOTPEQRITVDG 526
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 434 ANFPESKPTPGI-EPILQ-----TPRTVGL- DP 463
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 527 FSEFLSKVVRPIRIRRVYDQLAHIDLTLAQVAKNLGIELTDDQLNITPRPVNGLKKP 586
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 464 INQNETFTVPLFVIRPKG 480
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 587 -----SLSLVAIPDG 596
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 5
US-09-815-242-5469
Sequence 5469, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl U.
APPLICANT: Zysek, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21

```

; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5469
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5469

Query Match 3.9%; Score 101; DB 10; Length 398;
Best Local Similarity 20.3%; Pred. No. 0.15;
Matches 95; Conservative 62; Mismatches 158; Indels 154; Gaps 23;

QY 42 QAPLPTLTQHTTSGANDTILPLNNIQGDLVGMKKQKRFVFFQVNDATSFKTALKTVV 101
DB 53 QPGITTTPTQKT-----CNFVALDLKSKDRDAIKAMFKKWTVMADRMMDGDTVGKTSNNPLM 108
QY 102 PORITSAAILISDPSOQPLAFVNLGFSNTGLQALGITDLDGDA-----QFPGQGFADAAAN 156
DB 109 PPVDTGESIGLG-ASKLTIITF---GISKSLMKKIGLSSKIPDAFKDLPHFPNDQLIDDYS 164
QY 157 LGDQLSQWAPFTGTTIHGVFLIGSDODDFLDQFTDDISSTFGSSITQVQALSGSARPPD 216
DB 165 DGDIMIQACNSDQSVSHAV-----HNLVRFPRDIVKVRWSQS-----202
QY 217 QAGHEHFGFLDGISQPSVTCWETVPPGOAVVPPGIIILTRDGDGTGRPS-----266
DB 203 -----GFI-----SAKGET-----PRNLMAFKDGTINPRKSNQLKDYYFID 239
QY 267 --WALDGSFMAFHFOQKVPFENAYTLANAIPANSAGNLTOQEGAEFLGARMFGRWK-SG 323
DB 240 DGNWAKHGTVCVVRRIQIHETWDRTA-----LEEQEA-----FGRKRHS 280
QY 324 APIDLAPTADDPALGADPQNNNFYSDTLTDETRCFGAHVKTNPQDLGGPVDTFHA 383
DB 281 APLTGKKEFDEIDLKA---KDSHGEY---IIDK-----DAHTR-----LAKEANT-SI 321
QY 384 MRSSIPGTPETSDAELASGVTAQ-DRGLLFVEYQSIIGNGFRFQOINWANNANFPFSKPI 442
DB 322 LRRAFNVDGTDNR-----TGNFETGLLFIAFOKATQ-----OFIDIQNN-----361
QY 443 TPGIEPIIGTTPRTVGLDPLNQNETFTVPLFVIRKGEYFF-LPSIS 490
DB 362 -----LGSNDKLINE-----YITHRGSASFLVLPGVS 387

RESULT 6
US-09-815-242-12418
; Sequence 12418, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12418
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12418

Query Match 3.9%; Score 101; DB 10; Length 409;
Best Local Similarity 20.3%; Pred. No. 0.15;
Matches 95; Conservative 62; Mismatches 158; Indels 154; Gaps 23;
QY 42 QAPLPTLTQHTTSGANDTILPLNNIQGDLVGMKKQKRFVFFQVNDATSFKTALKTVV 101
DB 64 QPGITTTPTQKT-----CNFVALDLKSKDRDAIKAMFKKWTVMADRMMDGDTVGKTSNNPLM 119
QY 102 PORITSAAILISDPSOQPLAFVNLGFSNTGLQALGITDLDGDA-----QFPGQGFADAAAN 156
DB 120 PPVDTGESIGLG-ASKLTIITF---GISKSLMKKIGLSSKIPDAFKDLPHFPNDQLIDDYS 175
QY 157 LGDQLSQWAPFTGTTIHGVFLIGSDODDFLDQFTDDISSTFGSSITQVQALSGSARPPD 216
DB 176 DGDIMIQACNSDQSVSHAV-----HNLVRFPRDIVKVRWSQS-----213
QY 217 QAGHEHFGFLDGISQPSVTCWETVPPGOAVVPPGIIILTRDGDGTGRPS-----266
DB 214 -----GFI-----SAKGET-----PRNLMAFKDGTINPRKSNQLKDYYFID 250
QY 267 --WALDGSFMAFHFOQKVPFENAYTLANAIPANSAGNLTOQEGAEFLGARMFGRWK-SG 323
DB 251 DGNWAKHGTVCVVRRIQIHETWDRTA-----LEEQEA-----FGRKRHS 291
QY 324 APIDLAPTADDPALGADPQNNNFYSDTLTDETRCFGAHVKTNPQDLGGPVDTFHA 383
DB 292 APLTGKKEFDEIDLKA---KDSHGEY---IIDK-----DAHTR-----LAKEANT-SI 332
QY 384 MRSSIPGTPETSDAELASGVTAQ-DRGLLFVEYQSIIGNGFRFQOINWANNANFPFSKPI 442
DB 333 LRRAFNVDGTDNR-----TGNFETGLLFIAFOKATQ-----OFIDIQNN-----372
QY 443 TPGIEPIIGTTPRTVGLDPLNQNETFTVPLFVIRKGEYFF-LPSIS 490
DB 373 -----LGSNDKLINE-----YITHRGSASFLVLPGVS 398

RESULT 7
US-09-815-242-5898
; Sequence 5898, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.


```

APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5898
LENGTH: 1349
TYPE: prt
ORGANISM: Staphylococcus aureus
US-09-815-242-5898

```

Query Match 3.7%; Score 96; DB 10; Length 1349;
Best Local Similarity 20.9%; Pred No. 2.5; Indels 82;
Matches 57; Conservative 34; Mismatches 100; Gaps 12

QY 37 TNFPCOALPLTTHHTSSGNDILPLANNIGDILVGMKKQKE-RVFFQVNDATESFT 95

Db 803 TNKQK-----QDSTEKGI:SGTVTLVKNBNGEVLQTTKTKDKQKQKQFGLNGE- YKV 854

QY 96 ALKT---YVFORITSAAILISDPQCOPLAFNVLGFSNTGLDALGITDDLDGAPFPDQFA 152

Db 855 EFEPFSGTPTPOVGG-----TDEGIDNSGTSITTVIXDKNDITIDSGFYK 900

QY 153 DAANLGDLSQW-----VAPFTGTHGVFLIGSDODDEPL-----DQFTDDI 194

Db 901 PTYNLGDV--WEDTNKNGVDKDKBKGI:SGTVTLVKNBNDKVLKTVTTDNGKKQKQFIDLN 958

QY 195 SSTFESSITQVQALSGARPPDQAGHEHFGFLDGISPSVTGMETTVPCQAVVPPGII 254

Db 959 NGTV-----KYEFETPSGTPTSVTSNGDTEKDSNGLTTTGI- 996

QY 255 TGRDGDGTGRPSMALDSSFMFRHPOQKVPPEFN 287

Db 997 --KQADNMNT--LDSGF-----YTPKYS 1015

RESULT 8
 US-09-815-242-13137
 Sequence 13137, Application US/09815242
 Patent No. US20020061569A1
 GENERAL INFORMATION:
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Karl L.
 APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes in
 TITLE OF INVENTION: Prokaryotes
 FILE REFERENCE: ELITRA. 01A
 CURRENT APPLICATION NUMBER: US/09/815,242
 CURRENT FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21

```

?
? PRIOR APPLICATION NUMBER: 60/206,848
? PRIOR FILING DATE: 2000-05-23
? PRIOR APPLICATION NUMBER: 60/207,727
? PRIOR FILING DATE: 2000-05-26
? PRIOR APPLICATION NUMBER: 60/242,578
? PRIOR FILING DATE: 2000-10-23
? PRIOR APPLICATION NUMBER: 60/253,625
? PRIOR FILING DATE: 2000-11-27
? PRIOR APPLICATION NUMBER: 60/257,931
? PRIOR FILING DATE: 2000-12-22
? PRIOR APPLICATION NUMBER: 60/269,308
? PRIOR FILING DATE: 2001-02-16
? NUMBER OF SEQ ID NOS: 1410
? SOFTWARE: FASTSEQ for Windows Versio
? SEQ ID NO 13137
? LENGTH: 1349
? TYPE: prt
? ORGANISM: Staphylococcus aureus
US-09-815-242-13137

```

Query Match	3.7%	Score 96	DB 10	Length 1349
Best Local Similarity	20.9%	Pred. No.	2.5	
Matches	57	Conservative	34	Mismatches 100; Indels 82; Gaps 12;

```

QY      37  TNFPCQALPLRTHTHTSSCANDTILPLNNIOGILVGMKKQKE-RVFPQVNDATSEFT 95
Db      803  TNKKCK-----QDSTFKGISGVTVLTKNENGVLTQTTKDQCKQKQFGLNGLT- YKV 854
QY      96  ALKT--YVFORITSAAILISDPSCQPLAFVNLGFSNTGLQALGITDLDLDAQFPDQCF 152
Db      855  EFETPSGVTLPQVSGS-----TDEGIDBNGTSTTGVIKDKNDRTIDSGFYK 900
QY      153  DAANLGDLSQW-----VAPFTGTHGVFLGSDQDPL-----DQFTDI 194
Db      901  PTYNLGDVY--WEDTNKKGVDDKREKGISGVTVLTKENEDVLKTVTTDNGKQKQFIDL 958
QY      195  SSTFSSITVOVALSGSARPPDQAGHEHFGFLDGISQPSVTGNETTVPCQAVVPPGII 254
Db      959  NGTY-----KYEFETPSGVTPLSVTSGNDTEKDSNGLTTGVI- 996
QY      255  TGRDGDGTGRPSWALDGSFMAFRHPOQVPPFN 287
Db      997  --KDADNMT--LDSGF-----YKTPKYS 1015

```

```

RESULT 9
/ US-09-815-242-5803
/ Sequence 5803, Application US/09815242
/ Patent No. US20020061569A1
/ GENERAL INFORMATION:
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Kari L.
/ APPLICANT: Zyskind, Judith W.
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John D.
/ APPLICANT: Carr, Grant J.
/ APPLICANT: Yamamoto, Robert T.
/ APPLICANT: Xu, H. Howard
/ TITLE OF INVENTION: Identification of Essential Genes in
/ TITLE OF INVENTION: Prokaryotes
/ FILE REFERENCE: ELITRA.011A
/ CURRENT APPLICATION NUMBER: US/09/815,242
/ CURRENT FILING DATE: 2001-03-21
/ PRIOR APPLICATION NUMBER: 60/1191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27

```

```
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5803
; LENGTH: 991
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5803

Query Match
Best Local Similarity 22.1%; Score 95; DB 10; Length 991;
Matches 98; Conservative 48; Mismatches 184; Indels 114; Gaps 19;

QY 7 VVSVAVLVGSSSHVNAKLGAROTRTPL--TNFPG---QAPLTLTQHTTSGANDTI 61
DB 145 VASKLTDGSSVVVEKANEQFNKTVTRALLEEANKAGLTIEENVPIKINTNAGYSADKA 204
QY 62 LP-----LNNIQ-----GDILVGMKKQKE-RVFFQVNDATS 92
DB 205 LPKINDFANRIVYLNHHQADLDKYASDFRQSGSYEGDILDASEKIKRINGAIPOLNEKAK 264
QY 93 FKTALTYVUPORTISAAILISD--PSQOPLAFVNLGFSNTGL-QALGITDD-----LGDAQ 145
DB 265 LILALNNYMP-KIEKALNFAADVPAPQPKINQGLNIASQGIQDQANGQLNDKAGFVTQVR 323
QY 146 FDPQGFADAAANLGGD--SOWVAPFTGTHGVFLIGSDQDFDOFTDDISSTFG--- 199
DB 324 SRVGDITQDAIRRAQDLNRRNQOIP-----QNSAANNETSNSAPAAGNV 368
QY 200 -----SSITQVQALSGSARPPDQAGHEHFGFLDGISQPSVTGWETTVPFG 244
DB 369 ASTPPSAPSQDTAPNNVNTQNTAPNSNAPVSTTPQSTGKKGQSFVDIT---TT--- 421
QY 245 QAVVPGCIILTRDGDGTGRPSWALDGSFMAFRHFOOKVPEFNAYTLANAIPANSAGNLT 304
DB 422 QVSTANENTQNIITDKQVKSMEA-ALTGSLLSLN-----NLDTOAKA----- 462
QY 305 QQSGAFLGARMGRKSGAPIDLAFTADDPALGADPQRNNFDYSDTLTDTETRCPPGAH 364
DB 463 AQKDSQALRNISYGILASDKSPSFRSLDNVKSGLGYTTQYNOQFIDTLKE----- 513
QY 365 VRKTNPRDGLGSPVDTFHMRSSI 388
DB 514 IEK-NEUVDSLKSIDKVKAANNRI 536

RESULT 10
US-10-007-693-65
; Sequence 65, Application US/10007693
; Patent No. US20020146776A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT
; TITLE OF INVENTION: AND DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.515C2
; CURRENT APPLICATION NUMBER: US/10/007,693
; CURRENT FILING DATE: 2001-12-05
; NUMBER OF SEQ ID NOS: 157
; SEQ ID NO 65
; LENGTH: 978
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-10-007-693-65

Query Match
Best Local Similarity 3.6%; Score 92.5; DB 12; Length 978;
Matches 105; Conservative 47; Mismatches 213; Indels 141; Gaps 23;

QY 9 SVAVLVGSSSHVNAKLGAROTRTPLLTNFPQAPLPLTQHTTSGANDTILPLNNIQ 68

; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5803
; LENGTH: 991
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5803

Query Match
Best Local Similarity 22.1%; Score 95; DB 10; Length 991;
Matches 98; Conservative 48; Mismatches 184; Indels 114; Gaps 19;

QY 7 VVSVAVLVGSSSHVNAKLGAROTRTPL--TNFPG---QAPLTLTQHTTSGANDTI 61
DB 145 VASKLTDGSSVVVEKANEQFNKTVTRALLEEANKAGLTIEENVPIKINTNAGYSADKA 204
QY 62 LP-----LNNIQ-----GDILVGMKKQKE-RVFFQVNDATS 92
DB 205 LPKINDFANRIVYLNHHQADLDKYASDFRQSGSYEGDILDASEKIKRINGAIPOLNEKAK 264
QY 93 FKTALTYVUPORTISAAILISD--PSQOPLAFVNLGFSNTGL-QALGITDD-----LGDAQ 145
DB 265 LILALNNYMP-KIEKALNFAADVPAPQPKINQGLNIASQGIQDQANGQLNDKAGFVTQVR 323
QY 146 FDPQGFADAAANLGGD--SOWVAPFTGTHGVFLIGSDQDFDOFTDDISSTFG--- 199
DB 324 SRVGDITQDAIRRAQDLNRRNQOIP-----QNSAANNETSNSAPAAGNV 368
QY 200 -----SSITQVQALSGSARPPDQAGHEHFGFLDGISQPSVTGWETTVPFG 244
DB 369 ASTPPSAPSQDTAPNNVNTQNTAPNSNAPVSTTPQSTGKKGQSFVDIT---TT--- 421
QY 245 QAVVPGCIILTRDGDGTGRPSWALDGSFMAFRHFOOKVPEFNAYTLANAIPANSAGNLT 304
DB 422 QVSTANENTQNIITDKQVKSMEA-ALTGSLLSLN-----NLDTOAKA----- 462
QY 305 QQSGAFLGARMGRKSGAPIDLAFTADDPALGADPQRNNFDYSDTLTDTETRCPPGAH 364
DB 463 AQKDSQALRNISYGILASDKSPSFRSLDNVKSGLGYTTQYNOQFIDTLKE----- 513
QY 365 VRKTNPRDGLGSPVDTFHMRSSI 388
DB 514 IEK-NEUVDSLKSIDKVKAANNRI 536

RESULT 11
US-09-919-935-2
; Sequence 2, Application US/09919935
; Patent No. US20020049305A1
; GENERAL INFORMATION:
; APPLICANT: BATHE, BRIGITTE
; APPLICANT: MCKEL, BETTINA
; APPLICANT: PFEFFERLE, WALTER
; APPLICANT: HUTHMACHER, KLAUS
; APPLICANT: BINDER, MICHAEL
; APPLICANT: GREISSINGER, DIETER
; APPLICANT: THIERBACH, GEORG
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES WHICH CODE FOR THE METP GENE
; FILE REFERENCE: 211712US0X
; CURRENT APPLICATION NUMBER: US/09/919,935
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: DE 10053942.4
; PRIOR FILING DATE: 2000-08-02
; PRIOR APPLICATION NUMBER: DE 10109686.0
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/294,279
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-919-935-2

Query Match
Best Local Similarity 3.5%; Score 92; DB 10; Length 349;
Matches 45; Conservative 24; Mismatches 73; Indels 54; Gaps 7;

QY 2 DLSLFVSVAVLVGSSSHVNAKLGAROTRTPLLTNFPQAPLPLTQHTTSGANDTI 61
DB 76 DLGASPVSVTYGAGSGSTRERTSRI-ARRLAKQPLTT-----LVHLTLVNHTR----- 122
```

```
QY 62 LPLNNIGDILVGMKKQKQKRFVFOVNDATSFYKLTATYVPORTISAAILISDPSQOPLA 121
Db 123 -----EMKAILREYLELGLTNLALNGDPFGDELIG 152
QY 122 -FVN-----LGSNTGLQALGITD-----DLGDAOPFDQOPADANLGDLSQWVAPFTGT 171
Db 153 DWSTDDGLNASELIDIKSTPEFRERFDLGIASFPEGHFR-AKTLEBDTKYTLAKNGG 211
QY 172 TIHGVFLIGSDQDDPEL 187
Db 212 AFSYIQMFVEDYL 227

RESULT 12
US-09-764-870-313
; Sequence 313, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ14
; CURRENT APPLICATION NUMBER: US/09/764,870
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 313
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-764-870-313
```

```
Query March 3.5%; Score 91; DB 10; Length 707;
Best Local Similarity 20.3%; Pred. No. 2.8; Indels 172; Gaps 26;
Matches 96; Conservative 63; Mismatches 143;

QY 9 SVAVLGSSSHVNAKLGAROTRTPLLTNFPQAPLPTLTQHTTEGANDTILPLNNIQ 68
Db 237 SVKDVLEASHIE-QRGGETRRTA-----FGIEFARSEAPQKGRKAKKMYIITDGE 289
QY 69 GDILVGMKK---QKRFVFPFOVNDATSFYKLTATYVPORTISAAILISDPSQO 118
Db 290 SHDSFDLEKVIQOSER-----DNVTRYAVAVLGYNNRGINPETFLEIKYIASDPDX 343
QY 119 PLAFNLGFSNTGLQAL-GITDDLGDAQFP-DGQFADANLGDLSQWVAPFTGTTH- 174
Db 344 HF-----FNVYDEALADIDVALGDRIFSLGKTNKNETSFGLMSQ-----TGFSSHV 392
QY 175 --GVFL--IGSD-----ODDFLDQFTDISSTFGSSITQVQALSGSA 212
Db 393 EDGVILGAVGAVDMNGAVLKETSAQKVIPLRESYLKEPPEELK----- 436
QY 213 RPFDOAGHEHFGFLDGISQPSVTGMETTVFPGQAVVPRGIIILTGDDGTGTRPSWALDGS 272
Db 437 -----HGAYL-----GYTVT-----SVWSSRQG----- 454
QY 273 FMARFHFOQKPEFN-----AYTLANALPANSAGNLQOEGAEFLGAMFGRMKSGAPI 326
Db 455 ----RVYVAGAPRFNHTGKVLIFTMHN-----NRSULTIHQAMGQOIGS-YFSGSEITSYDI 505
QY 327 DLAPFADPALGAD-----PQRNNPFDYSDTLTDE---TRCPFG---AHV 365
Db 506 DGDGTVTLVLGAPMYFNGERGRKVVYELFQNLFYVNGILKDHSHONARFGSSIASV 565
QY 366 RKTNRPOD-----LGGVD-----TFHMRSSIPYGPB--TSDDELASGV 403
Db 566 RDLN--QDSYDVNVVVGAPFLSDHAGAIYFHFGRFSILKTPFORITASELATGL 617

RESULT 13
US-09-852-053-3
; Sequence 3, Application US/09852053
```

```
; Patent No. US20020055141A1
; GENERAL INFORMATION:
; APPLICANT: BERENS, STEPHAN
; APPLICANT: KALINOWSKI, JORN
; APPLICANT: PUBLER, ALFRED
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM STRAIN WITH
; FILE REFERENCE: MAS/21123/280248
; CURRENT APPLICATION NUMBER: US/09/852,053
; CURRENT FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: ERO 00110021.3
; PRIOR FILING DATE: 2000-05-12
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 3
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: PROPEP
; LOCATION: (1)..(637)
; OTHER INFORMATION: secD
US-09-852-053-3

Query March 3.5%; Score 90; DB 10; Length 637;
Best Local Similarity 18.6%; Pred. No. 3;
Matches 72; Conservative 62; Mismatches 118; Indels 136; Gaps 19;

QY 9 SVAVLGSSSHVNAKLGAROTRTPLLTNFPQAPLPTLTQHTTEGANDTILPLNNIQ 68
Db 97 TLVITVPGENTRAQOSLG-----QTSOLLFRVVGQAGMPDMT-----TLMP----- 137
QY 69 GDILVGMKKQKQKRFVFP-----QVNDATSFYKLTATYVPORTISAAILI 112
Db 138 -----ELFEMANRVEVGVITTEQANSLSEMTAVASTIABEGEATEPPTVSATPM 192
QY 113 SDPSQOPLAFVNLGFSNTGLQALGITDDLGDAQFPDQGFAD-----AANLGDLSQWVA 166
Db 193 DEPA-----NSIEATQRQOETIMLRTDGQSTDPYQVIAASSLMQCTTBMD 239
QY 167 PFTGT-----TIHGVFLIGSDQDDPELDFDFTDISST--FGSSITQVQALSGSA 212
Db 240 PLAGTDPRLPLVACDPAVGGVYL--DPAPLNGETDEENGARLTGNEIDTNRPITGCF 297
QY 213 RPFDOAGHEH--FGFLDGISQPSVTGMETTVFPGQAVVPRGIIILTGDDGTGTRPSWALD 270
Db 298 NA--QSGGMEISFAFKSG-----DGEESG--ATW--- 322
QY 271 GSFMAFRFHFOQKV-----PEFNAYTLANALPANSAGNLQOEGAEFLGAMFGRM 320
Db 323 -SLSLQYLQOQIAITLDSQVISAAPVIOATPVSATSITDFTQTEQDILANNLRIG-- 379
QY 321 KSGAPIDIA-----PTADDPALGA 339
Db 380 --ALPLSPAGENGERRGTTTVPSPSLGA 405

RESULT 14
US-09-779-081-2
; Sequence 2, Application US/09779081
; Patent No. US20020102270A1
; GENERAL INFORMATION:
; APPLICANT: Murdin, Andrew
; TITLE OF INVENTION: CHLAMYDIA ANTIGENS AND CORRESPONDING DNA FRAGMENTS AND
; FILE REFERENCE: 19721-009
; CURRENT APPLICATION NUMBER: US/09/779,081
; CURRENT FILING DATE: 2001-02-08
; PRIOR APPLICATION NUMBER: 09/427,533
; PRIOR FILING DATE: 1999-10-26
; PRIOR APPLICATION NUMBER: 60/106,046
; PRIOR FILING DATE: 1998-10-28
; PRIOR APPLICATION NUMBER: 60/132,271
```


JOURNAL Patent: JP 2000245468-A 1 12-SEP-2000;
 MEIJI SEIKA KAISHA LTD
 OS Geotrichum candidum Dec 1 (FERM P-15348)
 PN JP 2000245468-A/1
 PD 12-SEP-2000
 PF 26-FEB-1999 JP 1999050562
 PR MAKOTO SHODA, YASUFUMI KANNO, HIDEOTOSHI KUBOTA
 PC C12N15/09, C03F1/00, C07K14/37, C12N1/15, C12N1/19, PC
 C12N1/21, C12N5/10,
 PC C12N9/08/(C12N15/09, C12R1:645), (C12N9/08, C12R1:645),
 C12N15/00, PC C12N5/00,
 PC (C12N15/00, C12R1:645)
 CC
 FH Key Location/Qualifiers
 FT source 1. .1494
 FT /organism="Geotrichum candidum Dec 1 (FERM FT
 P-15348)".
 FEATURES Location/Qualifiers
 source 1. .1494
 /organism="unidentified"
 /db_xref="taxon:32644"
 BASE COUNT 310 a 458 c 408 g 318 t
 ORIGIN
 Query Match 100.0%; Score 1494; DB 6; Length 1494;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1494; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATCGCTGTTCGCTGTTGCGTTCGTTGCGTACTCGTGGTCCGAGCTCGCATGTC 60
 DB 1 ATCGCTGTTCGCTGTTGCGTTCGTTGCGTACTCGTGGTCCGAGCTCGCATGTC 60
 QY 61 AATGCTGCTAAACTCGGCGGAGACAGACGCTACGACACCCCTCTCACTAATTTTCG 120
 DB 61 AATGCTGCTAAACTCGGCGGAGACAGACGCTACGACACCCCTCTCACTAATTTTCG 120
 QY 121 GGAACAGCCCGCTGCGACTTAAACGACGATACGATGAGAGCGGGGCAACGATACA 180
 DB 121 GGAACAGCCCGCTGCGACTTAAACGACGATACGATGAGAGCGGGGCAACGATACA 180
 QY 181 ATTCTGCCCTGAACACATACAAAGGCGACATTTTGGTTCGATGAAGAACAGAGAA 240
 DB 181 ATTCTGCCCTGAACACATACAAAGGCGACATTTTGGTTCGATGAAGAACAGAGAA 240
 QY 241 CGCTTCGCTCTTTTCCAAAGTCAATGACGCAACCTCGTTCAAGACGCGTTGAAGACCTAC 300
 DB 241 CGCTTCGCTCTTTTCCAAAGTCAATGACGCAACCTCGTTCAAGACGCGTTGAAGACCTAC 300
 QY 301 GTGCTGAGCGCATACGTCGGCGGGGATTTTGAATTCAGATCCTTCTCAGACGCGTTG 360
 DB 301 GTGCTGAGCGCATACGTCGGCGGGGATTTTGAATTCAGATCCTTCTCAGACGCGTTG 360
 QY 361 GCTTTCGTAACCTCGGTTTCGACACAGGCTCCAGCGGCTTGGAAATTCAGCAT 420
 DB 361 GCTTTCGTAACCTCGGTTTCGACACAGGCTCCAGCGGCTTGGAAATTCAGCAT 420
 QY 421 CTGGGTGATCCACAATTCAGATGCTGATGTCGACAGCCGCAACCTCGGGGACGAC 480
 DB 421 CTGGGTGATCCACAATTCAGATGCTGATGTCGACAGCCGCAACCTCGGGGACGAC 480
 QY 481 CTGAGCAATGGGTGGCGCTTTTACTGTTACCAACATCCATGCTGCTTCTGATTTGGT 540
 DB 481 CTGAGCAATGGGTGGCGCTTTTACTGTTACCAACATCCATGCTGCTTCTGATTTGGT 540
 QY 541 AGCAGCAGGACGACTTCTTGGATCAGTTCCAGGATGATATCTCTCGACCTTTGGTTCC 600
 DB 541 AGCAGCAGGACGACTTCTTGGATCAGTTCCAGGATGATATCTCTCGACCTTTGGTTCC 600
 QY 601 TCCATCACTCAGGTGAGGCGCTCAGTGGTCTGCGGCTCCAGGATCAGGTGCTCAT 660
 DB 601 TCCATCACTCAGGTGAGGCGCTCAGTGGTCTGCGGCTCCAGGATCAGGTGCTCAT 660

QY 661 GAACACTTCGGGTTCTTCGACGGCATCTCGACGCCCTCAGTCACAGGCTGGAGAGACC 720
 DB 661 GAACACTTCGGGTTCTTCGACGGCATCTCGACGCCCTCAGTCACAGGCTGGAGAGACC 720
 QY 721 GTCTTCCCTGGACAGCGGTCGTCCACCTCGAAATTTATCTTCACTGGACGCGATGGGAC 780
 DB 721 GTCTTCCCTGGACAGCGGTCGTCCACCTCGAAATTTATCTTCACTGGACGCGATGGGAC 780
 QY 781 ACGGACCCGACCGTCGTGGGCTCTAGATGGAGTTTCATGGCATTCGGGACATTCAG 840
 DB 781 ACGGACCCGACCGTCGTGGGCTCTAGATGGAGTTTCATGGCATTCGGGACATTCAG 840
 QY 841 CAGAAGTCCCGAATTCAACGCTACAGCTCGCCAAACGCGATACCCGGAACAGCGG 900
 DB 841 CAGAAGTCCCGAATTCAACGCTACAGCTCGCCAAACGCGATACCCGGAACAGCGG 900
 QY 901 GGAACCTCACCACGAGGAGGTGTCAGAGTTCTTCGGCGCGCATGTTTCGGCCGTTGG 960
 DB 901 GGAACCTCACCACGAGGAGGTGTCAGAGTTCTTCGGCGCGCATGTTTCGGCCGTTGG 960
 QY 961 AAGAGCGGCGCGGATTCGCTCGCGCGGAGCGGCGGAGCGGCGGCGGCGGAC 1020
 DB 961 AAGAGCGGCGCGGATTCGCTCGCGCGGAGCGGCGGAGCGGCGGCGGCGGAC 1020
 QY 1021 CCGCAGAGAACAAATTTTCGATTTACTCAGACAGCTCGGCGGAGCGGCGGCGGCGG 1080
 DB 1021 CCGCAGAGAACAAATTTTCGATTTACTCAGACAGCTCGGCGGAGCGGCGGCGGCGG 1080
 QY 1081 TTCTGCTGACACGCTGAGGAGGAGGAACTTCGACAGGAGCTTCGGTGGAGCGGTCGAC 1140
 DB 1081 TTCTGCTGACACGCTGAGGAGGAGGAACTTCGACAGGAGCTTCGGTGGAGCGGTCGAC 1140
 QY 1141 TTCCACGCTATGCGGTTCAGATTCGCTGAGGCGGAGGAGGAGGAGGAGGAGGAG 1200
 DB 1141 TTCCACGCTATGCGGTTCAGATTCGCTGAGGCGGAGGAGGAGGAGGAGGAGGAG 1200
 QY 1201 TCGGGGCTGACTCGGCAAGACCGGCTCTCTTTTGGTTCGAGTACGAGTCCATTTGGT 1260
 DB 1201 TCGGGGCTGACTCGGCAAGACCGGCTCTCTTTTGGTTCGAGTACGAGTCCATTTGGT 1260
 QY 1261 AATGGGTTTCAGGTTTCAGGAGATTAATTCGGGAGGAACTTCGAACTTCCTCTTCTCAA 1320
 DB 1261 AATGGGTTTCAGGTTTCAGGAGATTAATTCGGGAGGAACTTCGAACTTCCTCTTCTCAA 1320
 QY 1321 CCGATCACGCTTGAATTCAGCTATTCATCGGCGGAGGAGGAGGAGGAGGAGGAGGAG 1380
 DB 1321 CCGATCACGCTTGAATTCAGCTATTCATCGGCGGAGGAGGAGGAGGAGGAGGAGGAG 1380
 QY 1381 CTCGACCCCTCAACAGGATGAGAGTTTCAGATACCGCTGTTTGGTTCGAGTCCGAGGAG 1440
 DB 1381 CTCGACCCCTCAACAGGATGAGAGTTTCAGATACCGCTGTTTGGTTCGAGTCCGAGGAG 1440
 QY 1441 GGGGAATCTTTTCTTTCGCTCTCTCTCTGCTCAGTCCGAGTATCGCTGCT 1494
 DB 1441 GGGGAATCTTTTCTTTCGCTCTCTCTCTGCTCAGTCCGAGTATCGCTGCT 1494

RESULT 2
 AB013135
 LOCUS 1586 bp mRNA linear PLN 08-MAY-1999
 DEFINITION Galactomyces geotrichum mRNA for Dyp, complete cds.
 ACCESSION AB013135
 VERSION AB013135.1 GI:4760439
 KEYWORDS Dyp.
 SOURCE Galactomyces geotrichum (strain:Dec 1) cDNA to mRNA.
 ORGANISM Galactomyces geotrichum
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Dipodascaceae; Galactomyces.
 1 (sites)
 SUGANO, Y., SASAKI, K. and SHODA, M.
 Cloning and sequence analysis of a novel decolorizing enzyme,
 peroxidase gene dyp from Geotrichum candidum Dec 1.
 Unpublished JOURNAL

REFERENCE 2 (bases 1 to 1586)
AUTHORS Sugano, Y., Sasaki, K. and Shoda, M.
TITLE Direct Submission
JOURNAL Submitted (21-APR-1998) Yasuichi Sugano, Tokyo Institute of

polyA_site

BASE COUNT	331 a	486 c	425 g	344 t
/gene="dyp"				
/note="13 a nucleotides"				

Query Match	100.0%;	Score 1494;	DB 8;	Length 1586;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1494;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0

[illegible]

LOCUS	PSU77073	1798 bp	mRNA	linear	PLN 06-JUN-1997
DEFINITION	Polypotracheae sp. peroxidase (cbop21) mRNA, complete cds.				
ACCESSION	U77073				
Db	504	CTCAGCCAAATGGGTGCGCCCTTTTAACTGGTACACACATCCATGAGTGTCTTTCTGATTTGGT	563		
Qy	541	AGCGACGAGACGACATCTTGGATCAAGTTCACGAGATATCTCTTGCACCTTGTGTTCC	600		
Db	564	AGCGACGAGACGACATCTTGGATCAAGTTCACGAGATATCTCTTGCACCTTGTGTTCC	623		
Qy	601	TCCATCACTCAAGTGTGACAGGCGCTCAGTGGGCTGTGGGCTCCAGAGATCAGGCTGTGTCAT	660		
Db	624	TCCATCACTCAAGTGTGACAGGCGCTCAGTGGGCTGTGGGCTCCAGAGATCAGGCTGTGTCAT	683		
Qy	661	GAACACTTCGGGTTCTGTGACGGCATCTCGAGCCCTAGTCACAGGCTGGAGACGAC	720		
Db	684	GAACACTTCGGGTTCTGTGACGGCATCTCGAGCCCTAGTCACAGGCTGGAGACGAC	743		
Qy	721	GTCCTTCCTGTGACAGGCGGTCGTCCACCTTGAATTAATCTCACTGACGCGATGGGAC	780		
Db	744	GTCCTTCCTGTGACAGGCGGTCGTCCACCTTGAATTAATCTCACTGACGCGATGGGAC	803		
Qy	781	ACGGGACCCCGACCGTCTGGGCTCTAGATGGAGTTTCATGGCATTCGGCATCTTCCAG	840		
Db	804	ACGGGACCCCGACCGTCTGGGCTCTAGATGGAGTTTCATGGCATTCGGCATCTTCCAG	863		
Qy	841	CAGAAAGTCCCCGAAATTAACGCGTACACGCTCCGCAACGGGATATCCCGGCAACAGCGCG	900		
Db	864	CAGAAAGTCCCCGAAATTAACGCGTACACGCTCCGCAACGGGATATCCCGGCAACAGCGCG	923		
Qy	901	GGAACACTCACCGACGAGAGGTGACAGTTCTCTGGCGCGCCCATTTTGGCCGTGTGG	960		
Db	924	GGAACACTCACCGACGAGAGGTGACAGTTCTCTGGCGCGCCCATTTTGGCCGTGTGG	983		
Qy	961	AAGACCGGCGCGCGCATTTGACTCTGCGCCGACGCGCGACGACCCAGCGCTCGCGCCGAC	1020		
Db	984	AAGACCGGCGCGCGCATTTGACTCTGCGCCGACGCGCGACGACCCAGCGCTCGCGCCGAC	1043		
Qy	1021	CCGCGAGAGAACACAAATTTTGATTACTCAGACACAGCTGACGAGACGAGACCCCGCTGCCCC	1080		
Db	1044	CCGCGAGAGAACACAAATTTTGATTACTCAGACACAGCTGACGAGACGAGACCCCGCTGCCCC	1103		
Qy	1081	TTCGGTGACACAGTGAAGAACGAAACCCCTGCACAGACCTTGGGTGACCGGTGCACACC	1140		
Db	1104	TTCGGTGACACAGTGAAGAACGAAACCCCTGCACAGACCTTGGGTGACCGGTGCACACC	1163		
Qy	1141	TTTCAACGCTATGCGGTCCAGTATCCCGTACGCGCCGAGAAAGTCTGATGACAAATTTCG	1200		
Db	1164	TTTCAACGCTATGCGGTCCAGTATCCCGTACGCGCCGAGAAAGTCTGATGACAAATTTCG	1223		
Qy	1201	TGCGGCGTGAAGTGTGCGCAAGACCGCGGCTCTTTTTCGTGAGTACAGTCCATTAATTGGT	1260		
Db	1224	TGCGGCGTGAAGTGTGCGCAAGACCGCGGCTCTTTTTCGTGAGTACAGTCCATTAATTGGT	1283		
Qy	1261	AATGGGTTCAAGTTCTCAGACAGATTAACTGGGCGAAACAATGGGAATTCCTTTCGCAAA	1320		
Db	1284	AATGGGTTCAAGTTCTCAGACAGATTAACTGGGCGAAACAATGGGAATTCCTTTCGCAAA	1343		
Qy	1321	CCGATCAGCGCTGGAATTGAGCTATCATCGGCCAGACGATCCACGCACTGTGCGCGGG	1380		
Db	1344	CCGATCAGCGCTGGAATTGAGCTATCATCGGCCAGACGATCCACGCACTGTGCGCGGG	1403		
Qy	1381	CTTCAACCCCTCAACAGATGAGCGTTTCAAGTACCGCTGTTTGTATCCGAAAGGC	1440		
Db	1404	CTTCAACCCCTCAACAGATGAGCGTTTCAAGTACCGCTGTTTGTATCCGAAAGGC	1463		
Qy	1441	GGGGAATATCTTTTCTTGGCCCTTATCTCTGCGCTCACTGGAGCATATGCTGCT	1494		
Db	1464	GGGGAATATCTTTTCTTGGCCCTTATCTCTGCGCTCACTGGAGCATATGCTGCT	1517		

U77073.1 GI:2160704
Polyporaceae sp.
Polyporaceae sp.
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
Aphyllophorales; Polyporaceae.
1 (bases 1 to 1798)
Han, Y.-H.
Direct Submission
Submitted (04-NOV-1996) Microbiology, Seoul National University,
San 56-1, Shillim-dong, Kwanak-gu, Seoul 151-742, Korea
Location/Qualifiers
1..1798
/organism="Polyporaceae sp."
/db_xref="taxon:54443"
1..1798
/gene="cpop21"
1..1500
/gene="cpop21"
/notes="oxidizes phenolic lignin model compounds"
/codon_start=1
/product="peroxidase"
/protein_id="AAB58908.1"
/db_xref="GI:2160705"
/translation="MOLKHFLLAATAFSAVYHVKRARSPLIGSFQCPPLP
TIAEVSSAGNDSLFPENIQDILVGMKDKXKFFVFFHNNATFKSLKYAPANI
TSVQTIIGPASGQPOAFVNLAFSHTGFGALGVADDDLDQTAFTAGQPADAPSDGDTST
WEEAFKTVNVDGVLGSDVTTNQYRDDLAKLGDWTVLLDLSAARPGAEGHGE
HFGLDGIISNPTIPGFTFPQAVVDGVIIFAGRTNDPVTNRPSWALDGSFLVFRKL
KOLVPEFHKWTLONLQNOAGNLTVEEGALLIGSRMFGNWSGAPIDLPDVIDDPTLG
NPPORNDPNYIHPGEDLITDTRCPPTAHVRKTNPRDLLEAOGLIPDLFHAIRACTPY
VIGQVDAESNNTSIDRGLAVEYOSVINSNGFRPQQLAWANNANFPFNKSEPLGLDLP
GTEGTRQTFGLDPRNASLTIPIQIIISNGGEYFSPSITALVEHFGA"
BASE COUNT 420 a 487 c 423 g 468 t
ORIGIN

Query Match 26.8%; Score 399.8; DB 8; Length 1798;
Best Local Similarity 58.8%; Pred. No. 9e-79;
Matches 834; Conservative 0; Mismatches 552; Indels 33; Gaps 7;
QY 79 GCAGAGCAGCGGTACGACACCCCTCCTCACTAAATTTCCGGGACAAAGCCCGCTGCCG 138
DB 73 GTGAAGGGGCAAGATCGTCTCCCTCATTTGGCTCATTTCCCGGTACGCCACCACTCCCT 132
QY 139 ACTCTAACGAGCATACGATCAGAGCGGGGCCAAAGATACAAATCTGCCCTGAAACAC 198
DB 133 ACTATCGCGAGGTTCACTCTCTCTCGCGGAATGATAGCCTTCCGTTCGAGAATA-- 190
QY 199 ATACAGGCCACATTTTGGTTGGCATGAAGAAACAGAGAAACGCTTCGCTTTTCCAA 258
DB 191 -TCCAGGGAGATATTCTGTGGGCATGAAGAGGACAAAGGAGAAGTTGTCTTCTTCCAC 249
QY 259 GTCATGACCAACCTCGTTCAAGACGGCTTGAAGACCTACGTCCTGAGCGCATCACG 318
DB 250 ATTAAACATCGGACCAAGTTCAAGAGTTCTTGAGACGTACGCTCCCGCAATACACT 309
QY 319 TCGCGCGCGATTTTGATTCATGATTCCTCAGCAGCGCTTGGCTTTCGTCACACCTCGG 378
DB 310 TCTGTCCAAACATCATCGGTCTCGCTCGGTCAGCCCCAGCGGTTTGTGAACCTTGGC 369
QY 379 TTTTCGAACACAGCCCTCCAGGCGCTTGGAAATTACCGACCATCTGGGTGATGACAAATTC 438
DB 370 TTCTCCCATACCGGTTTCGATGCTTGGGTGTGGCCATGATTTACAAAGATACATGCTTTC 429
QY 439 CCAGATGCTCAGTTCGCAGACGCGCAAAACCTCGGGACACCTCAGCCCAATGGGTGGCG 498
DB 430 ACAGCAGGACAGTTTGGCAGCGCCCTCTTTTGGCGACGACACCACTTGGGAGGAA 499
QY 499 CTTTATCTGGTACCAACCATCTGATGCTTTCTGATAGCGACCAAGGACGACTTC 558
DB 490 GCTTTCAGGGGCCAACACGTTGATGGGTTTCTTGTGATTTGGTAGCGAGTGTCAACACC 549
QY 559 TTGATCAGTTTCCAGGATGATATCTCTTCGACCTTTGGTTCCTCCATCACTCAGGTGCAG 618

Db 550 ACAAACAATACCGGATGACCTCAAGGCTAAGCTCGGTGACGCTTGGACCGCTTGTCTC 609
QY 619 GCCTCAGTGGGTGCTCGGCTCCAGGAGATCAGGCTGTGATGAACATCTCGGGTTCCTC 678
Db 610 GATCTTTGACAGTGCAGCAAGCGCGGGCTGAGAAGGACACGAACATTTGGCTACTTG 669
QY 679 GACGGCATCTCGAGCGCTCAGTCACAGGCTGGGAGACACCGTCTTCCCTGGACAGGCG 738
Db 670 GACGGTATCTCCAATCTCTATATCCCGGCTTCGGTACACCC---ATTCCCTGGCCAAGCC 726
QY 739 GTCGTCCCACTCGAATTTCTCTACTGACGCGATGGGACACGCGCAC---CCGACCG 795
Db 727 GTAGTGGATTCCGCTGTATCTTCGCGCGCGCACCAACGATCCCTCACAACCGTCT 786
QY 796 TCGTGGCTCTAGATGGGAGTTTCATGGCATTTCCGGCATTTCCAGCAGAGGTCCCCGAA 855
Db 787 TCGTGGCTCTGGATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 846
QY 856 TTCAACGGGTACAGCTCGCAACGCGATACCGCGAAACAGCGGGGAAACCTTACCCAG 915
Db 847 TTCCATAAATGGACACTGGACAAACGCTCTGC---AGAAATCAGGCTTGGAAACCTTACAGTC 903
QY 916 CAGGAAGGTGCAGAGTTCTCTCGGCGCGCATGTTCCGGCGCTTGGAGAGAGCGCGCGCG 975
Db 904 GAGGAGGGGCTCTACTTTTGGCTCTCGATGTTTGGCTCGTGGAAACAGTGGCGGACCA 953
QY 976 ATTGACCTCGCGCGAGCGGACGACCCAGCGCTCGCGCGCCGACCCGAGAGGAAACAC 1035
Db 964 ATCGATCTCACTCTCAGCTCGACGATCTCTACCTTTGGCAATGATCTCTCAGCGCAACAC 1023
QY 1036 AATTCGATTTAC-----TCAGACACGCTCAGCGACGAGACGCGCTGCCCTTC 1083
Db 1024 GACTTCAACTACATACATCCAGGAGAAGACCTTACGACAGACGAAACCGCTCCCGCTTC 1083
QY 1084 GGTGCACAGTGTAGGAAGACGAAACCTCG-----ACAGACCTCGGTGGACCGTCGAC 1137
Db 1084 ACTGCACAGTTTCGCAAGACCAACCCCGGTGATTTAGAGACCAAGGTCTTATCCCTGAT 1143
QY 1138 ACCTTCCACGCTATGCGGTCCAGTATCCCGTACGGCCAGAAACGTCCTGATGAGAACTT 1197
Db 1144 CTGTTCCATGCCATCCGTCGGGAAACGCTTATGGGCTTACGATACCGATCCGATCC 1203
QY 1198 GCGTCGGGCTGATCTCGCAAGACCGGCTCTTCTTTTCTGTCGAGTACAGTCCCATATT 1257
Db 1204 AACAGCAACACTACTTTCGATCGATCGTGTCTGGCATTTTGTGGAGTACAGTCCGCTATT 1263
QY 1258 GGTAAATGGGTTTCAAGTTCAGCAGAGTTAACTGGGCGAAACAATCGAACTTCCCTTCTCC 1317
Db 1264 TCTAATGGATTCCGATTCCAGCAGCTCAATTTGGGCCAACAACCGAAATTTCCCTTCAAC 1323
QY 1318 AAACCGATCACGCTCGAATTTGAGCCTATCATCGGCGCAGACGACTCCAGCACTGTCCGC 1377
Db 1324 AAGTCTGAGCCACTCGGACTTGTATCCGTTATCGGACAAAGTACTCGCCAGACATTCGCG 1383
QY 1378 GGGCTCGACCCCTCAACAGAAATGAGAGTTCACAGTACCGCTGTGTGTGATCCGGAAG 1437
Db 1384 ---TTGGACCCCGGAAACGCTCGGATAGTCTCAGCATTTCTCAGATAATCATCTCTAAC 1440
QY 1438 GCGGGGAATACTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1476
Db 1441 GGAGGGAATATTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1479

RESULT 4
AF468656
LOCUS AF468656 1735 bp mRNA linear PLN 02-MAY-2002
DEFINITION Termitomyces albuminosus peroxidase TAP mRNA, complete cds.
ACCESSION AF468656
VERSION AF468656.1 GI:20386143
KEYWORDS Termitomyces albuminosus.
SOURCE Termitomyces albuminosus
ORGANISM Termitomyces albuminosus

Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
Agaricales; Tricholomataceae; Termitomyces.
1 (bases 1 to 1735)
Johjima, T., Ohkuma, M. and Kudo, T.
Novel peroxidase from Termitomyces albuminosus
JOURNAL
AUTHORS
REFERENCE
TITLE
JOURNAL
1 (bases 1 to 1735)
Johjima, T., Ohkuma, M. and Kudo, T.
Direct Submission
Submitted (14-JAN-2002) JST Biorecycle project, Kasetsart Research
and Development Institute, Bangkok 10900, Thailand
Location/Qualifiers
1. 1735
/organism="Termitomyces albuminosus"
/strain="ATCC42010"
/db_xref="ATCC:42010"
/db_xref="taxon:79931"
34. 1548
/codon_start=1
/product="peroxidase TAP"
/protein_id="AA021606.1"
/translation="MQLKFLAATAFAVALSQPIAAVHYKRASSPLIYPPGQALP
/TAAQVSTSGADDSIPRENTQADILVEMKQNEFYFPHNNAATFESVLKTYAPANI
TSAVTLIGPISAPOAVVNVAFSGAGALVDNLGDTAFGTGQFADSDGDDTN
TWEPAFKGTNDGVFLIISDQSIITQYODDLQKLDADWTVVYDLGSAARPGDFKH
EHEGYLDGINSPLIKGFGEPLPGQAFIDPGLILVGRNDVTVTPPAALGFLAFPK
LKDLPFRHYTIDNALQNSLSTEEGSLTSGRMRGMSQNLIDLPDVPDL
KDPNRRNNRYNTHPEDEPRTDQSRCPFTAHIRKTRNDLESQNLIDEPFLHRAQPT
YGPENVAESSNSTQIDRGIAVEYQSNISNGRFQVWMANNKRPFKSISEPL
DPVIGDKGNRIITGTLNPLNVSLSLIPDIVSNGEYFAPSITSIVETLAA"

BASE COUNT 437 a 483 c 371 g 444 t
ORIGIN

Query Match 21.8%; Score 325.6; DB 8; Length 1735;
Best Local Similarity 56.2%; Pred. No. 3.2e-62;
Matches 811; Conservative 0; Mismatches 589; Indels 42; Gaps 9;

85 CAGACGGCTAGACACCCCTCCTCACTAATTTCCGGGACAAAGCCCGCTCGCATTA 144
112 CGGGCAAGTATATCCCCCTATGTCCTCTTCCGGGCGAGGGGCACTACCAACTGCA 171
145 ACGGAGCATACGACTGAGAGGGGGCAAGCATACATTTGCCCTGAGAACATACAA 204
172 GCCCAGTTCAATC--AACTCTGGCGGCATGACAGCATCCCTTGGAATAATTTAG 228
205 GCGCAGATTTGGTTGGCATGAGAAAGAGAAAGCGCTTGTCTTTTCCAACTCAAT 264
229 GCAGATATCTTGGGTGGATGAGAAAGAGAAAGAAATTTGTCTTTCACATCAAC 288
265 GACCAACCTGTTCAACAGCGGTTGAAAGCTACGCTGAGCGCATCACTCGCGC 324
289 AACGGGCAAGTTCAAAAGGTTCTCAAGACATATGCTCGGCTAATATTAATCCGTT 348
325 GCGATTTGATTTAGATCTTCTCAGCAGCGGTTGGCTTTCGTAACCTCGGGTTTCG 384
349 GCTACCTTAATAGTCCATCTCCGCCCAACCAAGCAGATGTAAGTGGCTTTCTCT 408
385 AACACAGGCTCCAGCGCTTGGAATTAACGAGATCTGGGTGATGACAAATCCAGAT 444
409 CAGGCTGGGTGGGTGACATGGGTGATGATTAATCTCGGGAACATGCTTACTGGG 468
445 GGTGATTCGAGAGCGCGCAA---CCTCGGGGACGACCTTCAGCAATGGGTGGCGCT 501
469 GGGCAGTTGCGTGAATGCTCGATGCTTGGGAGATGATCAACACATGGGAGGCTGCT 528
502 TTACTGTACACCATCATGATGCTTTCTGATTTGATGGAGCAGAGACGATCTCTTG 561
529 TTCAAGGACCAATATGAGCGGCTTTTCTGATCATTAATGAGCAACAGATTTCAATT 588
562 GATCATTTCAAGGATGATATCTTTCGACCTTGTGCTTCTTCATCACTCAGGTGAGCG 621
589 ACTCAGTACAGGATGATCTCCAAAGCAAGTAGAGAGCGGTGACTGTGTATACAT 648

622 CTCAGTGGGCTCGGCGCTCCAGAGATCAGCTGTGTCATGAAACATTTGGGTTCTTGAC 681
649 TTGAGTGGCCAGACAGACCTGGGACCAATTTGTTGATCAGAACTTTGGCTATTTGAC 708
682 GGCATCTCGAGCGCTTCATGACAGAGCTGGGAGACAGCGCTTCTTCTGACAGCGCTC 741
709 GGTATCTCTAACCCTTCATATAAAGGATTTGGGAAAC---CTTGCTCGCCAGGCTTT 765
742 GTCCACCTGAATTAATCTCAGTACGAGGATGGGAGACGGGAC---CGACGCTG 798
766 ATTGATCGGATATCATTTAGTAGGCGCGCAATGATACGACACTCGCCCGCG 825
799 TGGGCTTAGAGAGATTTATGAGCATATCCGACATTCAGAGAAAGTCCCGAATTC 858
826 TGGGCTTAGAGATTTATTTCTGCGCTTCGCAAGCTCAAAAGCTTTGTTCCGAATTT 885
859 AACCGTACACGCTCGCAACCGGATACCCGCAACGCGGGGAAACCTCACCCAGCAG 918
886 CACAAATATAGCTAGACAAATGCTGCA---AACCAATCAGGAAACCTTCAACCGAA 942
919 GAAGTGCAGAGTTCTCGGCGCGCGCATGTTGGCCCTTGGAAAGCGCGCGCATT 978
943 GAAAGTCTCTTCTTTGGGATCAAGGATGTTGGTGTGGAACAGCGGTGACCAATC 1002
979 GACCTCGCGCGAGCGGAGACACCCAGCGCTCGCGCGCGAGAGGAAACAAAT 1038
1003 GATTTAACCCCGCATGCTGATGATCCGCGCTCTCGGACCAACCTTAACGAAACAAAT 1062
1039 TTGATTA-----CTCAGACACCTGACGAGACGAGCGCTCGCTTGGT 1086
1063 TTCAACTATATCATCCCGGAAAGATCTGCGACACAGACGCGCTTCTTTCAGC 1122
1087 GCACCGTGAAGAAAGAACCC-----TCGACAGACCTGGGTGGAACCGGTGACAC 1140
1123 GCGCAATTCGCAAGACCAATCCCGGATCTGGAATCTAGAAATTCATCCCGAATTC 1182
1141 TTCACGCTATGCGGTCAGATCCCGTACGAGCCGAGAAACGTGTGATGAGAACTTGC 1200
1183 TTCAACGCTATGCTGAGAGAAACCTTATGACACAGATATCTATGCCAGTCCAGC 1242
1201 TCGGCGGTGACTGCGAAGACCGCGGCTTCTTTCTGCTGAGTACGATCATTATGCT 1260
1243 TCTAATGACACCAATATGATCGCGGTCTGCGATTTGTGATGATCAATCAATATCA 1302
1261 AATGGTTCAAGTTCCGACGATTAATCTGGCGAACAATGGAATCTCTTTTCCAAA 1320
1303 AACGATTTTGGCTTCCAAAGGTCATTTGGGCTAATAACAACAAAGTTCCCTTCATAG 1362
1321 CCGATCAC---GCCTGGAATTGACCTATCATGCGGACAGCA-----CTCCAGGACT 1371
1363 AGTATCTGAGAGCGGGGCTTGACCTGTCTATGGAAGATTAAGGTGCAACCGCATTT 1422
1372 GTGCGCGGCTCGACCCCTCAACCAAGATGAGAGCTTCAAGTACCGCTTTGTGATC 1431
1423 ACAACCGGCTTGAATCCCTCAAGCTTTGAGAGAGTTGAGCATTTCCGAACTTATGCTC 1482
1432 CCGAAGGCGGGGAAATCTTTTCTTGCCCTTATCTCTGCGCTCACTCGCATATGCT 1491
1483 TCCATGCGCGGTGAATCTTCTGCTCTTCTATCACTATTTGAGAGACTTGTGCC 1542
1492 GC 1493
1543 GC 1544

RESULT 5
AF429315 125020 bp DNA linear PRI 18-JAN-2002
LOCUS Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
DEFINITION
ACCESSION AF429315
VERSION AF429315.1 GI:17646244
KEYWORDS

Disease-Like 2 (HDL2)"
 repeat_region complement (35581..35746)
 /rpt_type=tandem
 /rpt_unit=cig
 complement (<36507..>36887)
 /gene="JPH3"
 /note="JPH3"
 mRNA complement (<36507..>36887)
 /gene="JPH3"
 /product="junctional protein 3"
 complement (<36507..>36887)
 /gene="JPH3"
 /note="component of the junctional complex between plasma
 membrane and endoplasmic reticulum"
 /codon_start=1
 /product="junctional protein 3"
 /protein_id="P41409.1"
 /db_xref="GI:17646245"
 /translation="MSSGRFNPDDGSSYCGMEDGKAHGVCTGPKGGGYTGSWS
 HGFELGVYTPSGNTYQGTMAQGRKIGLEKSKVYKGEWTHGFKGKGYVRECAG
 NGAKYEGTNGSLDGYETYSYG"
 BASE COUNT 29056 a 32731 c 30696 g 28283 t 4254 others
 ORIGIN
 Query Match 3.7%; Score 54.6; DB 9; Length 125020;
 Best Local Similarity 11.7%; Pred. No. 0.098;
 Matches 99; Conservative 363; Mismatches 366; Indels 18; Gaps 3;

QY 204 AGGCGACATTTGGTGGCAGTAAGAAAGAGAGAGCGCTTCTTTCCAACTCA 263
 DB 17778 RGGTGYMKKKGGGKSSSKMKSSSTRRRGSAKSCSTMWGMCMSCMSMAKSYM 17719
 QY 264 TGACCAACCTCGTTCAAGACGCGCTTGAAGACCTACGCTGCGATCAGCGC 323
 DB 17718 CYCTMYKMSYMYKYCYSCMWSSTSYSCCKMWSMCSYCKMYYGMSYSISM 17659
 QY 324 GCGCATTTGATTCAGATCTCTCAGACAGCCGTTGCTTCAACTCGGGTTTC 383
 DB 17658 GYVSTSCKKYKCSWSMYKCKTSKSYRRKSYWYGKAKKKYVCAARRRMSYK 17599
 QY 384 GAACACAGCGCTCCAGGCGCTTGAATTAACGACATTTGGGTGACAACTCCCA 443
 DB 17598 CAKMWSYCYMSYCTMYTYSCTYKSTCYKRGYVGSYKTCYSAGKSRSMYCMWRS 17539
 QY 444 TGGTCAGTTCGACAGCGCGCAAACTG-----GGAGACCTCAGCAATGGCG 496
 DB 17538 SKSSWSMWSARSSWCMGWRKRSKSAWGMWSKGRSTGMKRAKSKTGSTGS 17479
 QY 497 CGCCTTTTACGTATCAACCATTCATGCTGTTCTTGTGATTTGAGCAGCAGCA 556
 DB 17478 TGRMVKKKKYSKYSRGMKKKTCYCMKYKRYKRTSCMWMYMKSGYRKRCCMK 17419
 QY 557 TCTTGATCACTTCAAGATATATCTTTGACCTTTGGTTCCTCAT-----C 606
 DB 17418 KGCTGYRMSKSKSGYKMRGSSYSTSCMSGWSMCMKMYKKRRSMRSGM 17359
 QY 607 ACTCAGGCGAGGCGGCTGAGTGTCTCGGCTCAGAGATCAAGGCTGATCAACA 666
 DB 17358 SSKMRGAGAGCTSSSMWSRKRKSKYSYKYGKMKMGWMRGSKYWSMKM 17299
 QY 667 TTCGGTTCCTGACGCGATCTCGAGCCCTCAGTACAG-GCTGGAGACAGCGCTT 725
 DB 17298 KRSSMSKYSYKYSRGMKGRSTKSAKSMRWAGSKCTYGSYSNMNRNRMG 17239
 QY 726 CCTGACAGCGGCTGCTCCCACTGGAATTATCTCACTGAGCGGATGGGACAG 785
 DB 17238 KTGCAVYWRBRAMNNGMAAGCTTCCCANITGGGGAAGGCGSASRASCYKR 17179
 QY 786 CACCGACCGTGGGCTTCAAGATGGAGTTTCAATGCGGCACTCCAGCAAA 845
 DB 17178 MSSSCTRSRGRRCMSKCRGRSGKSMGMRGSGSTISASGSGGCTCYCMGWRK 17119
 QY 846 GGTCCCGAATTAACGCGTACGCTGCGCAACCGATACCCGGAACGCGCGGAA 905

Db 17118 GRKMSRKKMKMYKSYRRKRWMTCMKICYSMYAMCRSMCKMCKSCCGCSMGMS 17059
 QY 906 CCTCACCAGAGAGAGTGCAGAGTTCCTCGGCGCGGATGTTGGCGGTGAGAG 965
 DB 17058 YSYGKYSWGMKSYMYKYSKYSTKAMRSKRGMGTGATGGAGSYGCGGSG 16999
 QY 966 CGGCGCCCGATGACCTCCGCGCAGCGGAGAGCAGCCAGCGCTCGCGCCGCA 1025
 DB 16998 SMGKGKSKTRGSSAGKSSYKMKSCCARYKMSYKCYGRMWMCMGYYCWMGCMSSC 16939
 QY 1026 GAGGAA 1031
 DB 16938 WWSMR 16933
 RESULT 7
 AE001862
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 1 (bases 1 to 213732)
 White, O., Eisen, J. A., Heidelberg, J. F., Hickey, E. K., Peterson, J. D.,
 Dodson, R. J., Haft, D. H., Gwinn, M. L., Nelson, W. C., Richardson, D. L.,
 Mofatt, K. S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,
 Vamathevan, J. J., Lam, P., McDonald, L., Utterback, T., Zalewski, C.,
 Makarova, K. S., Aravind, L., Daly, M. J., Fraser, C. M., et al.
 Genome sequence of the radiotolerant bacterium *Deinococcus*
radiodurans R1
 Science 286 (5444), 1571-1577 (1999)
 2 (bases 1 to 213732)
 White, O., Eisen, J. A., Heidelberg, J. F., Hickey, E. K., Peterson, J. D.,
 Dodson, R. J., Haft, D. H., Gwinn, M. L., Nelson, W. C., Richardson, D. L.,
 Mofatt, K. S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,
 Vamathevan, J. J., Lam, P., McDonald, L., Utterback, T., Zalewski, C.,
 Makarova, K. S., Aravind, L., Daly, M. J., Fraser, C. M., et al.
 Submitted (02-NOV-1999) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 Location/Qualifiers
 1. 213732
 /organism="Deinococcus radiodurans"
 /strain="R1"
 /db_xref="taxon:1299"
 /chromosome="2"
 653..1435
 /gene="DRA0001"
 /gene="DRA0001"
 /note="similar to GB:D26185.SP:P37522 PID:467381
 PID:580906 GB:AL009126 percent identity: 54.96, identified
 by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="chromosome partitioning ATPase, putative, Para
 family"
 /protein_id="AF12301.1"
 /db_xref="GI:6460595"
 /translation="MYSAYKTLTVFNHAGAGCTSLTVNGYELANGGLRVLLDDP
 QANLTWMLGISGTYREMYVPAVADGPPLSPKAKGLDVIPIHVSILAAVEGGMKGV
 GAQGRRLAAVEAGDYVALIDSPPSLGLALIAALADOMIVPPTTRKGDALPG
 LQALTEYREVRLTVLVYVPTFYDARRRHDEVLADLKAHSLPLARVPQREAVVL

gene	gene
CDS	CDS
<p>DSTAQCAPUSEYAPGCTPVHADYQRLTADIAAIGVAYPCGENA" 1432..2313 /gene="DRA0002" 1432..2313 /gene="DRA0002" /note="similar to GB:L34077 GB:X85964 PID:407375 PID:757761 PID:1163135 percent identity: 57.34; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="chromosome partitioning protein, ParB family" /protein_id="AAFL2300.1" /db_xref="GI:6460594" /translation="MTRRRPERRRDLGLLGETPVDLSQANDIRALPVNELKVGSTOP RSLFOLRSELAESIRAHGVLOPLLVSDVQGYEIVAGERWRRAQAQAGLAEPVYV RSLNEQARAALINLQDNLDNVDEVDGKLEIALTLGLRERERAKLMQLLRAVP GHEHLDQOVFSMGETWRTFAKNKRLINLPQPVLEARAGLPLTLGSAVASAPPER QAEELKLAQNGARSQQLQALQTPSQTSAVTEHFAKVLSSKRFSLGLDTPREALDR WLARMPVRQAIDQES" complement(2418..3881) /gene="DRA0003" complement(2418..3881) /gene="DRA0003" /note="similar to GB:M88334 SP:P25526 PID:147901 GB:U00096 PID:1789015 percent identity: 69.60; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="succinate-semialdehyde dehydrogenase" /protein_id="AAFL2294.1" /db_xref="GI:6460588" /translation="MTTDLRTYSVTSRQAYFGCEWNRNAPRNFEVRHPNGEVI VADCTPTDARQIDAEALREWQNPYERKILRRWHDLMPEHKEELAQLMTLEM KPISTRGVHVAAGSIWCABEAGRIAGERINLRFPHKRGITISEPVGIYAVTPWN FPMGITRAAPALAAACGVMLKPAELSPMTALYTELWLKAGGPANTFOVLTPDAS ALTQPMNDSVRKLTFTGSTEVRLLYQAAQTIKRVSLLELGGHAPFLVFDADLER AASEVASFRNSGQTCVCTNRYVQGVAAEFIRLLTEKTAALQGLDFDFDQTQGP VVEQAGLQVQVODALTKGAQATGGQVSGSLFPQPTVLVDVAPDSLLIREETFGP VAPVTFIDTEEGLRANDEYGLAAYTRDLGRAFRIAGLEYGIVGINDGLPSSA APHVPFGGMKNSGVREGHGWGLEYLETKFVSLGLS" 4024..4668 /gene="DRA0004" 4024..4668 /gene="DRA0004" /note="similar to Gp:2924378 percent identity: 55.29; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAFL2299.1" /db_xref="GI:6460593" /translation="MNSLSLAALVILMAGRSTRMGOPKALLPLGGMPLCRHAAQTIVE AGYGEWAVLPPGETGDAVRAALQPLRWNTNPOPAQGLSSFOAAGALWAESAAP AVVFTLADNPLVRDTHOALGTVFGAEQPAALITRYGEVAAPTLRLRLRLRLEFF PSDHGPTLRLRELKIKIDRPAMELLDIDPEAWAQAQLFAEVTGQAQGP" complement(4719..5897) /gene="DRA0005" complement(4719..5897) /gene="DRA0005" /note="similar to GB:L33464 SP:P47734 PID:496118 percent identity: 74.23; identified by sequence similarity; putative" /transl_table=11 /product="alcohol dehydrogenase, zinc-containing" /protein_id="AAFL2293.1" /db_xref="GI:6460587" /translation="MKAIWQGVNKKVGVETVPDPTLLLPDIAIVKITSTAICGSDLHL LDGYTPSMETGIDILGHEFVGEVSEVRKLKVGDRVVVFNFIACGCDPCPKRGFFS ACDQNSPNRMAEALGSGGALFGYSHMYGQAGQAYRVFPADVGPFKISGL KDEQVLTLDITPTGYOAAEOCNI VPGRDVAVFGAGVPQFAARSQMLGAHVIV DRVPERLMAEAGCOTINYEKEDVLLSREATGGGPDHVIDAVGMEAHGHGPGSTM DKVQQLKLTDFRITALRWAILSCAKGGTVSNPGVYGGVLVDKMPGAFAKGLIFRMG</p>	
gene	gene
CDS	CDS
<p>QTHTHRHPSSLGRIEAGQIDPSFVITHRASLDQAPELYKTPRDKQCCIKVVLNPWA" complement(5894..6700) /gene="DRA0006" complement(5894..6700) /gene="DRA0006" /note="similar to PID:1653488 percent identity: 48.46; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAFL2292.1" /db_xref="GI:6460586" /translation="MDRRSSPOLRTMTWNREQAPHAGATGPNSSPGATMQPLERSL TGAVGAALAAAGLSRSGLKVKVLGGLAGLJIAAAGQNPLATKAEQNDEGETTV RMTADPGKRTAWRSLPGARIENSGEVLFRPAPGARGTEVVVRLTYRPPGGSAGAVIA ELFNOEPSQQLRDDLMPFKREGLGHAFQAQSSSRAALTPPNKRGDLV" 5839..7744 /gene="DRA0007" 5839..7744 /gene="DRA0007" /note="similar to GB:AL009126 percent identity: 62.98; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="ABC transporter, ATP-binding protein, putative" /protein_id="AAFL2291.1" /db_xref="GI:6460585" /translation="MTTFEQQVSKTYGHVTALSDNLTLTGELTALLGPNCGAGKSTA IGLLGLSAPSGQVRVLGADPRNDVPARIGAMPQESALPAGLTVREAVTLFASFPY APLGVDEALADLQGVAGRAAQQGGKRLAFALAVGDPPELLLIDEPITGMDAQ SRAAFVETGLRARGRTLLITTHYLEAERTADRVVVMNGRILADDTPOGLSGVG CARVFSQDLVQAELEIRPGVSQVAVQADAGRADLRTSVPEALLAALIGSTFTFSDLEV RRATUEEAYLQLTGQDMTAVTRSA" 7861..8652 /gene="DRA0008" 7861..8652 /gene="DRA0008" /note="similar to GB:AL009126 percent identity: 52.46; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAFL2290.1" /db_xref="GI:6460584" /translation="MTTLPNVPARTSALPVLQTLGELRRLRSPMFTVCAIGFPV MFFALFGLPAVQEGYATDPHVPVLTQFAAYSLSLALFSGRAVATERSGGWLRLL RSSPLPVLVYFVKTLAALAFGAVSLALLYAFAPHPAGVTPLGLALLALLKLLGMI PLVALGCLGFLASPQAACILANILSVVMSFASGLFVPLDQLPGVFQQLAPLLPAYHV SQIATNTVSGQTASEPAHMLAAFTLVFGTLAVMWGLKRDESREG" 8809..9819 /gene="DRA0009" 8809..9819 /gene="DRA0009" /note="similar to GB:AL009126 percent identity: 55.80; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAFL2289.1" /db_xref="GI:6460583" /translation="MLFWGTLVAFVAVYVFTFGKRDPDRWAYVQWGLSALLMVALGDM LYPSALSFLVYGGSLIGWQRWALGAAINALLMVGRLLTLTPERSQCDIFILLFI LASAVANHATVGLRARRRRLAQVLEKLAADARERIALRDHLGLHTLSVLKLS ELAGKLAEKHPQRAAEI REVERISREALSEVRAAVQYRGSGLAEARAKVALDAA GIRLVITTLPELADLSTLAMVLRVAVTNVVRHSGASELRLTRGQANGQLFIQD DGRGDADPEGTGLNSMRRLRAMGGTTFGRDGHGTGRLOAFVPLSAGTAGLKAPELLEK TS" 9816..10451 /gene="DRA0010" 9816..10451</p>	

Query Match 3.7%; Score 54.6; DB 1; Length 213732;
 Best Local Similarity 50.1%; Pred. No. 0.099;
 Matches 197; Conservative 0; Mismatches 184; Indels 12; Gaps 2;

QY 925 GCAGAGTTCCTCCGCGCGCGCATTTGGCCGTGGAAGAGCGCGCGCATTTGACCTC 984
 DB 144981 GCCGAGCTCTTGGCGCCAACTGGTCGAGCGTGGCGCGCGCGCATTTGACCTC 145040
 QY 985 GCGCGCAGCGCGCGAGCAGCAGCCGCTCGCGCGCGCGCAGAGAAACAATTTGCAT 1044
 DB 145041 GCTCCCAAGAGAGAGCAGCCCGAGCTGGGCCACAGACCCCAACAGATTTCACC 145100
 QY 1045 TACTCAGACAGCTGAGCAGAGCAGCGCGCTGCGCTTCGGTGCACAGTGAAGAAGC 1104
 DB 145101 TACAAAAACAGC---CGSAAAGGTTGGAAGTGGCGCTGGCGCATTCGCCCGCATG 145157
 QY 1105 AACCTTGACAGAGACCTGGGGTGGACCGGTGCACACTTCCAGCTATGCCGTCCAGATC 1164
 DB 145158 AACCCGGGTGACACCAAGCTGGAACTGCTCACCGAGTGAACATCCACCGGATTATTGCG 145217
 QY 1165 CCGTACGCGCCAGAAACGCTGATGACAGAACTTGCTGCGCGCGTGCACAGACCGC 1224
 DB 145218 CGCGCCACGGCGTATGGCCCGCCTATGACCCCAAGGCCGACAGCTGGCGGAAGACAG 145277
 QY 1225 GGTCTTCTTTTCGTGAGTACAGTCCATTAATTTGTAATGGTT-----CAGGTTTC 1275
 DB 145278 GTGCAAGCTGGCGCTGATCTTCATTCATGACGCCCAAGGAGTGAACACCGAGGTTTC 145337
 QY 1276 CAGCAGATTAACTGGCGGACAAATGCCAATTC 1308
 DB 145338 CTCGMAAAGATGATCAACAGCGCAACTTC 145370

RESULT 8
 LOCUS AV034474/c 2040 bp DNA linear BCT 10-JUN-2001
 DEFINITION Methylobacterium extorquens dioxygenase gene, complete cds.
 ACCESSION AV034474
 VERSION AV034474.1 GI:14334396
 KEYWORDS
 SOURCE Methylobacterium extorquens.
 ORGANISM Methylobacterium extorquens
 Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 Methylobacterium group; Methylobacterium.
 1 (bases 1 to 2040)
 Toyama, H., Matsushita, K., Adachi, O. and Lidstrom, M.E.
 Cloning and analysis of the dioxygenase gene in downstream of pqdFG
 unpublished
 2 (bases 1 to 2040)
 Toyama, H., Matsushita, K., Adachi, O. and Lidstrom, M.E.
 Direct Submission
 Submitted (11-MAY-2001) Biological Chemistry, Yamaguchi University,
 1677-1 Yoshida, Yamaguchi 753-8515, Japan
 Location/Qualifiers

FEATURES
 source
 1..2040
 /organism="Methylobacterium extorquens"
 /db_xref="taxon:408"
 1..933
 /note="orf88"
 /codon_start=1
 /transl_table=1
 /product="dioxygenase"
 /protein_id="AAK59392.1"
 /db_xref="GI:14334397"
 CDS
 1..933
 /note="orf88"
 /codon_start=1
 /transl_table=1
 /product="dioxygenase"
 /protein_id="AAK59392.1"
 /db_xref="GI:14334397"
 /translation="MQLTGLHVTATTAQADNLAFYTRVLGRLVKYTNODVSAY
 HLFVADGRASPGDITFDWPAERGTDSISRTLLRVGASFDWVRHLARQVS
 HHFLAIEDGLTLDFEDGEGORLSLVHGGVGAHPAGAPVPEHIGLGPRLTV
 SOPERFAVLTVLGFPRARFELPEGEVVFETGEGGPAAYOLKGTGAPRQAG
 AVHHVARRIDADYDAMDADLKORRVSSGPVRYRSLYFEPNGLIELFATDGG
 FADTETVESIGERIALPPEPRAEIENGLKPL"

BASE COUNT 319 a 704 c 674 g 343 t
 ORIGIN

Query Match 3.5%; Score 52; DB 1; Length 2040;
 Best Local Similarity 59.5%; Pred. No. 0.36;
 Matches 88; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 1208 TGACTGGCAAGCCCGGCTTTCTTTCTCTGAGTACAGTCCATTTATGTAATGGT 1267
 DB 1273 TGCCCAAGACGGGCTGCGACTCTGTCATGAGCTTCAAAATGCCCCACAGT 1214
 QY 1268 TCAGTTCCAGGATTAACTGGGCGAACAATGCAACTTCCCTTCCAAACCGATCA 1327
 DB 1213 TCGGTTTCAGCAGGAGCTTTGGCGCAACAACAACTTCCCTTCACGCCAGCCGC 1154
 QY 1328 CGCCTGGAATTGAGCCTATCATCGGCCA 1355
 DB 1153 CCATGAGTGCATCCGGTATCGGCCA 1126

RESULT 9
 SCF42 34713 bp DNA linear BCT 12-MAY-2002
 LOCUS Streptomyces coelicolor cosmid f42.
 DEFINITION AL137165 AL645882
 ACCESSION AL137165.2 GI:20520910
 VERSION
 KEYWORDS ABC transporter protein, ATP-binding component; Arac-family
 regulator; branched-chain amino acid ABC transport permease;
 branched-chain amino acid transport ATP-binding protein;
 DNA-binding protein; glycosyl hydrolase; glycosyl transferase;
 hydrolase; integral membrane protein; lipA; lipase; lipoprotein;
 lipR; membrane protein; oxidoreductase; transcriptional activator.
 Streptomyces coelicolor A3(2)
 Streptomyces coelicolor A3(2)
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 1 (bases 1 to 34713)
 Redendach, M., Klier, H.M., Denapate, D., Eichner, A., Cullum, J.,
 Kinashi, H. and Hopwood, D.A.
 A set of ordered cosmids and a detailed genetic and physical map
 for the 8 Mb Streptomyces coelicolor A3(2) chromosome
 Mol. Microbiol. 21 (1), 77-96 (1996)
 97000351
 MEDLINE
 PUBMED 8643436
 2 (bases 1 to 34713)
 Oliver, K. and Harris, D.
 Unpublished
 3 (bases 1 to 34713)
 Thomson, N.R., Parkhill, J., Barrell, B.G. and Rajandream, M.A.
 Direct Submission
 Submitted (24-JAN-2000) Streptomyces coelicolor sequencing project,
 Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
 CB10 1SA E-mail: barrel@sanger.ac.uk Cosmids supplied by Prof.
 David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
 Colney, Norwich, Norfolk NR4 7DH, UK
 May 9, 2002 this sequence version replaced gi:6752308.
 Notes:

Streptomyces coelicolor sequencing at The Sanger Centre is funded
 by the BBSRC and Biowulf Genomics
 Details of S. coelicolor sequencing at the Sanger Centre are
 available on the World Wide Web.
 (URL: http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are
 numbered using the following system eg SC7B7.01c. SC (S.
 coelicolor), 787 (cosmid name), 01 (first CDS), c (complementary
 strand).
 The more significant matches with motifs in the PROSITE database
 are also included but some of these may be fortuitous. The length
 in codons is given for each CDS.
 Usually the highest scoring match found by fasta -o is given for
 CDS which show significant similarity to other CDS in the database.
 The position of possible ribosome binding site sequences are given
 where these have been used to deduce the initiation codon. Gene
 prediction is based on positional base preference in codons using a
 specially developed Hidden Markov Model (Krogh et al., Nucleic
 Acids Research, 22(12):4768-4778(1994)) and the Frameplot program
 of Bibb et al., Gene 30:157-66(1984) as implemented at

<http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, ggg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid F42 Lies alongside and overlaps with cosmid F15 on the AseI-F genomic restriction fragment.

FEATURES

```

source      Location/Qualifiers
1. .34713
   /organism="Streptomyces coelicolor A3(2)"
   /strain="A3(2)"
   /db_xref="taxon:100226"
   /clone="Cosmid F42"

misc_feature 2. .892
   /notes="Pfam match to entry PF01315 Ald_Xan_dh_C, Aldehyde
   oxidase and xanthine dehydrogenase, C terminus, score
   -294.50, E-value 0.00012"
   /complement(1187..2644)
   /gene="SCF42.02c."
   /notes="SCO0692"
   /complement(1187..2644)
   /gene="SCF42.02c."
   /notes="SCF42.02c."
   /notes="SCF42.02c, possible membrane protein, len: 495 aa.
   Rich in the amino acid Ala. Contains possible membrane
   spanning hydrophobic domains."
   /codon_start=1
   /transl_table=1
   /product="putative membrane protein."
   /protein_id="CAB69664.1"
   /db_xref="GI:6752310"
   /db_xref="SPTREMBL:Q9L2L1"
   /translation="MTTSSSSRSPEYAIASLDLPTGTTTTTGTARPPTAEARAS
   SARPSQESDWDVDFDADLPEPEPDLFASGLPAPGPGALDADAGAPAAAG
   ADAAAHGVDGVLDVRAAADRPLDEVDVLTITLERSFQAQATVDALRAVGQVNR
   VEDVTLVGLLTPRQSDSADEAIRAAAECSVEDVTRLMALLHRTPLPHCGQAAV
   RAAATGRPELVELIGLAESCHMRPERPDARLRADDAEDDDGGRAPRSAGRG
   CTSAGADRGRERRRPPRRGRORDETAAARDRAARTABORPARAVRDRDKARA
   VDRRAARRAPGAPAWLTVAVLACGVAYFPLHRGDASAPAYGALGSLALCLVLA
   LLTVRPVPMIAAAAVGPAALAAAKLYGSPASRVPTVDTLTAPVWVAVAVAA
   SLVALMALCVRVASQDSRRRPARPVAVSSRTAAD"
   /RBS
   /gene 2816..2820
   /RBS 2822..3169
   /gene "SCF42.03"
   /notes="SCO0693"
   /RBS 2822..3169
   /gene "SCF42.03"
   /notes="SCF42.03, possible membrane protein, len: 115 aa.
   Contains a possible N-terminal signal sequence and a
   possible membrane spanning hydrophobic domain. Rich in Ala
   and Leu."
   /codon_start=1
   /transl_table=1
   /product="putative membrane protein."
   /protein_id="CAB69665.1"
   /db_xref="GI:6752311"
   /db_xref="SPTREMBL:Q9L2L0"
   /translation="MHRAPEEGAVDTTIVVAVLAAALVLAALVRLVRRRLD
   RRAGLTPGRWVFWGAVLYLVLPADLLPDPVYLDLIGVILLALRSARGATGGRLOPAE
   RRRTTGPGCDYAA"
   /RBS 3280..3283
   /gene 3289..3480
   /notes="SCO0694"
   /RBS 3289..3480
   /gene "SCF42.04"
   /notes="SCF42.04, unknown, len: 62 aa."
   /codon_start=1

```

```

/transl_table=1
/product="hypothetical protein SCF42.04."
/protein_id="CAB69666.1"
/db_xref="GI:6752312"
/db_xref="SPTREMBL:Q9L2K9"
/translation="MOPFALNVARPAVELEATTPVYVDSGLQLNLLDGRVAACDHAL
LRELGTSTAGSKTHFDD"
/RBS 3497..4483
/gene "SCF42.05"
/notes="SCF42.05, unknown, len: 328 aa; similar to
TR:Q9KZT0 (EMBL:AL33832) hypothetical protein SCE6.08
from Streptomyces coelicolor (326 aa) fasta scores; opt:
371, Z-score: 407.1, E(): 4.8e-15, 35.629% identity
(39.145% ungapped) in 334 aa overlap."
/codon_start=1
/transl_table=1
/product="Conserved hypothetical protein SCF42.05."
/protein_id="CAB69667.1"
/db_xref="GI:6752313"
/db_xref="SPTREMBL:Q9L2K8"
/translation="MTVLILTSEEDVTADMVVVHNLNAGVVPVRLDPADLTDSVALSG
EFAGSGFRGLSSGRLYSIGGLRSVWYRRPGGATRAEAPSALWTEAGQALYGLR
GSGARWNPDAARARYKQWLRLAQRCGLPVPATLITTPPRAAREFAERTPDLVVK
PVSGAHQDPLPLAVETSRVPEADFSAVAHGPTLLQRRVAKRADIRLTAVGEBELAAAR
KTALASLDPEVDVRFAGSGEPFWRPAEYVPRVAEGRAYLRAAGLAYGALDPAEDGDG
TWFLECNQSGQGFGEVDVTGQTIARTIAEWLARPCAADAVEPGGPDAAAVG"
/gene="SCF42.06c"
/notes="SCO0696"
/complement(4823..5317)
/gene="SCF42.06c"
/notes="SCF42.06c, possible integral membrane protein, len:
164 aa. Contains multiple possible membrane spanning
hydrophobic domains."
/transl_table=1
/product="putative integral membrane protein."
/protein_id="CAB69668.1"
/db_xref="GI:6752314"
/db_xref="SPTREMBL:Q9L2K7"
/translation="MTKILLSLVLAALVAVGPVTVVAASMPAAARRVPAATVGAD
GAAGDAGGVGTVRLHRCIKRYAGLGIAVPLGFATAAAGVLSDAWLVASITLTAV
AAGLLAAFLPRQELLEELADGREVERARTARLAMPFTGVFNLLWATVTVLIVRPGS
TTGA"
/RBS 5572..6459
/gene "SCF42.07"
/notes="SCF42.07, possible AraC-family regulator,
pseudogene, len: 309 aa. Similar to several other putative
AraC-family transcriptional regulators from Streptomyces
coelicolor e.g. TR:CAB61299 (EMBL:AL13299) SCF55.29 (311
aa) and TR:Q9XA73 (EMBL:AL096822) SCGD3.05 (334 aa).
Contains a frameshift after codon 188, following a
possible deletion event."
/codon_start=1
/transl_table=1
/product="putative AraC-family regulator (pseudogene)"
/complement(6487..6852)
/gene="SCF42.08c"
/notes="SCO0698"
/complement(6487..6852)
/gene="SCF42.08c"
/notes="SCF42.08c, unknown, len: 121 aa."
/transl_table=1
/product="hypothetical protein SCF42.08c."

```

Query Match 3.5%; Score 51.8; DB 1; Length 34713;
 Best Local Similarity 48.5%; Pred. No. 0.41;
 Matches 143; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 870 GCTGCGCAACGCGATACCCCGAAGACGCGGGAACCTTACCCAGACGAGAGAGTGCAGA 929
 Db 19596 GCTGCGCGCGCGCGCTGCTCGACGCGACCGGTCGATCGCCGACCTGCGCGCGCGCG 19537
 QY 930 GTTCTGCGCGCGCGCGATGTTGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 989
 Db 19536 GCGCTTGGCTGCGCTCGGAGCAGCGCGCGCTACGTGTCGTGCGCGCGCGCGCGCGCG 19477
 QY 990 GACGCGCGAGACCGACCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1049
 Db 19476 GCG 19417
 QY 1050 AGACACGCTGACGAGACGAGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1109
 Db 19416 CGGCGCGCGAGACCGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19357
 QY 1110 TCACAGAGCCTGCGTGGACCGCGTGCACCTTGCACGCTATGCGCGCGCGCGCGCG 1164
 Db 19356 GCACGCGCGGAGCGCGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19302

RESULT 10
 DROACT88F 2150 bp DNA linear INV 26-APR-1993
 DEFINITION D.melanogaster actin gene, complete cds, locus 88F.
 ACCESSION M18830.101065
 VERSION M18830.1 GI:156772
 KEYWORDS actin.
 SOURCE Drosophila melanogaster (Canton S) DNA.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 2150)
 Sanchez,F., Tobin,S.L., Rdees,U., Zulauf,E. and McCarthy,B.J.
 Two Drosophila actin genes in detail. Gene structure, protein
 structure and transcription during development
 J. Mol. Biol. 163 (4), 533-551 (1983)
 JOURNAL MEDLINE 83189087
 PUBMED 6405041
 COMMENT

[1] also shows the complete sequence of the actin gene at
 cytological locus 79B and found that the 79B and 88F actin genes
 exhibit a combination of co-ordinate and differential patterns of
 messenger accumulation. They appear to follow parallel patterns of
 transcription through embryogenesis to the prepupal stage but
 subsequently differ greatly with regard to the proportion of
 message from each gene in pupae. Of the Drosophila actin introns
 studied so far, the intervening sequences of the 79B and 88F genes
 are the only ones inserted at the same position.

FEATURES
 Location/Qualifiers

1..2150
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 prim_transcript 496..>1874
 /note="actin mRNA"
 CDS join(684..1608,1669..1874)
 /note="actin"
 /codon_start=1
 /protein_id="AAA28321.1"
 /db_xref="GI:156773"
 /translation="MCDADAGALVINDSGMCKAGFAGDAPAPVPSIVGRPHQGV
 MVMGQDSYVGDSEAGSKGILTLKYIEHGIITMDMEKIMHTFYNLRAPEBH
 PVLLTEALPNKAREKTOIFETFPSPAYVAIQVLVLSVSGTGLVSDGEV
 SHTVPIYEGEPAALIRLADGLDIDVIMKILTEGCTFTTAEREIVRDKEKIC
 YVALDPEEMATAASTLEKSYELPQGVITIGNERFRPEPALFQSEFGMSGCIH
 ETVYNSIMKCDVDIRKDLVANSVLSGGITVPGIADMQEITTLTAETIKIKIAP
 ERKYSWVIGLIIAASLTFOOMWISKOEYDESGPSIVHRKCF"
 <684..1608

/note="actin"
 /number=1
 1609..1668
 /note="actin intron"
 1669..>1874
 /note="actin"
 /number=2

BASE COUNT 544 a 562 c 517 g 527 t
 ORIGIN 77 bp upstream of HinfI site; locus 88F.

Query Match 3.1%; Score 46.2; DB 3; Length 2150;
 Best Local Similarity 48.3%; Pred. No. 7.2;
 Matches 129; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

QY 461 CCCGAACCTCGGGGAGCAGCTTCAGCAATGGGAGCGCTTTTACTGATGACCACTCC 520
 Db 1243 CCGATTACCTGATGAGATCTGACGAGCGCGGCTATACCTTACCAACCGCGGAGC 1302
 QY 521 ATGATGCTTTTGTATTGTGAGCAGCAGCATCTTTGTGATCAGTTACGAGATGATA 580
 Db 1303 GTGAGATCGTGGCGGACACCAAGAGAGAGCTGTGCTACGTGGCTTGAGACTTGAGAGACG 1362
 QY 581 TCTCTTGACCTTTGTTGTTCTTCATCACTAGTGACAGCGCTCAGTGAGTTCGCGCTC 640
 Db 1363 AGATGGCCACCGCTGCGCGCTCCACCTCGCTGAGAAGTGTAAGACTTCCTGACGCGCC 1422
 QY 641 CAGGAGATCAGCGTGGTCATGAAACACTTGGGTTCTCGACGCGCATCTGCAGCGCTCAG 700
 Db 1423 AGGTGATCACCATTGGCAAGAGGCTTCGCTGCGCGCGCGCTGTTTCAGCGCTCGT 1482
 QY 701 TCACAGGCTGGGAGACGACCGCTTCTC 727
 Db 1483 TCTCGGCGCATGAGTGTGCGGCATCC 1509

RESULT 11
 SCE66/c 28527 bp DNA linear BCT 12-MAY-2002
 LOCUS Streptomyces coelicolor cosmid B66.
 DEFINITION AL358989 AL645882
 ACCESSION AL358989.1 GI:8894754
 VERSION ABC transporter ATP-binding subunit; acetyltransferase;
 dehydratase; dimethyladenosine transferase; fatty acid desaturase;
 gale, UDP-glucose 4-epimerase; galk, galactokinase; galT,
 galactose-1-phosphate uridylyltransferase; isopentenyl
 monophosphate kinase; MarX-family; membrane; ppc,
 phosphoenolpyruvate carboxylase; prs, ribose-phosphate
 pyrophosphokinase; pch, peptidyl-cRNA hydrolase; ribosomal L25p
 family protein; secreted; sodium:solute symporter; TetR-family
 transcriptional regulator; trans-aconitate methyltransferase;
 two-component response regulator; UDP-N-acetylglucosamine
 pyrophosphorylase.

SOURCE Streptomyces coelicolor A3(2).
 ORGANISM Streptomyces coelicolor A3(2)
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1 (bases 1 to 28527)
 Redendach,M., Kiezer,H.M., Denapate,D., Eichner,A., Cullum,T.,
 Kinashi,H. and Hopwood,D.A.
 A set of ordered cosmids and a detailed genetic and physical map
 for the 8 Mb Streptomyces coelicolor A3(2) chromosome
 Mol. Microbiol. 21 (1), 77-96 (1996)
 JOURNAL MEDLINE 97000351
 PUBMED 8843436
 COMMENT 2 (bases 1 to 28527)
 REFERENCES Seeger,K.J. and Harris,D.
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 28527)
 AUTHORS Bentley,S.D., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
 TITLE Direct Submission
 JOURNAL Submitted (28-JUN-2000) Streptomyces coelicolor sequencing project,
 Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
 CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.

David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK

Notes:

Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics

Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web

(URL: http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are numbered using the following system eg SC7B7.01c, SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid E66 lies between and overlaps cosmids E41 and E87 on the AseI-E genomic restriction fragment.

FEATURES

source	1..28527	Location/Qualifiers
	/organism="Streptomyces coelicolor A3(2)"	
	/strain="A3(2)"	
	/db_xref="taxon:100226"	
	/clone="cosmid E66"	
misc_feature	1..107	/note="nominal overlap with Streptomyces coelicolor cosmid SCE66"
misc_feature	405..458	Nominal overlap with Streptomyces coelicolor cosmid E41"
	/notes="Pfam match to entry PF00132 hexapep, Bacterial transferase hexapeptide (four repeats), score 6.20, E-value 1.1e+02"	
misc_feature	459..512	/notes="Pfam match to entry PF00132 hexapep, Bacterial transferase hexapeptide (four repeats), score 8.40, E-value 55"
misc_feature	558..611	/notes="Pfam match to entry PF00132 hexapep, Bacterial transferase hexapeptide (four repeats), score 7.90, E-value 65"
misc_feature	663..716	/notes="Pfam match to entry PF00132 hexapep, Bacterial transferase hexapeptide (four repeats), score 3.80, E-value 2.3e+02"
misc_feature	843..896	/notes="Pfam match to entry PF00132 hexapep, Bacterial transferase hexapeptide (four repeats), score 6.80, E-value 93"
RBS	1153..1157	
gene	1166..2140	
	/genes="SCE66.02"	
	/notes="SCO3123; praA2"	
CDS	1166..2140	
	/genes="SCE66.02"	
	/notes="SCE66.02, praA2, ribose-phosphate pyrophosphokinase (EC 2.7.6.1), len: 324 aa; strongly similar to many eg.	

SW:PI4193 (KPRS BACSU) ribose-phosphate pyrophosphokinase from Bacillus subtilis (317 aa) fasta scores; opt: 1000, z-score: 1178.1, E(): 0, 46.8% identity in 314 aa overlap. Contains Pfam match to entry PF00156 Pribosyltran, Phosphoribosyl transferase domain and Prosite match to PS00103 Purine/pyrimidine phosphoribosyl transferases signature."

/codon_start=1

/transl_table=11

/product="ribose-phosphate pyrophosphokinase"

/protein_id="CAB95916.1"

/db_xref="GI:8894756"

/db_xref="SPTREMBL:O9K3U0"

/translation="MTGKTKTGEKKMFFSGRAHPELAEEVAQQLGVVPTKARDPFA NGEIYVVOESARGADCFLIQSHPTAPINKWNEQLIMIDALKRASARITVIVPYGY ARQDKHRRGEPISARLIADLMKTAGIRLILAVLDHTDQIQGFFGPDVHLFPALLLA DYVGAKRSLKLTIVVSPDAGRVRVADRCMLGAPLAIYHKRRDKVANQVTVHVVVG DVGKICVLVDDMDITGOTICAAADALPHAEDVITATHGVLSPAADRLKNSKVS ERFVNTLPSASELELDKITVLSIAPTITARAVREVFEDGSVTSLFDEQ"

1592..2008

/gene="SCE66.02"

/note="Pfam match to entry PF00156 Pribosyltran, Phosphoribosyl transferase domain, score 100.60, E-value 1e-26"

1838..1876

/gene="SCE66.02"

/note="PS00103 Purine/pyrimidine phosphoribosyl transferases signature"

2193..2252

/note="Score 52: 20/22 (90%) matches, 0 gaps, 16 base loop"

2343..2347

2353..2349

/gene="SCE66.03"

/note="SCO3124"

2353..2349

/gene="SCE66.03"

/note="SCE66.03, ribosomal L25p family protein, len: 198 aa; similar to many members of the ribosomal L25p family (includes ribosomal protein, general stress proteins and glutamyl-tRNA synthetases) eg. SW:PI4194 (CTC BACSU) general stress protein from Bacillus subtilis (703 aa) fasta scores; opt: 327, z-score: 393.9, E(): 1.8e-14, 31.6% identity in 196 aa overlap. Contains Pfam match to entry PF01386 Ribosomal L25p, Ribosomal L25p family and Prosite match to PS00017 ATP/GTP-binding site motif A (p-loop)."

/codon_start=1

/transl_table=11

/product="ribosomal L25p family protein"

/protein_id="CAB95917.1"

/db_xref="GI:8894757"

/db_xref="SPTREMBL:O9K3T9"

/translation="MSEYKLTAEFTFEGKGAARIRBDNKVPGVLYGHGSDPLHLTL PGHELLALRTSNVLIALDQGNELAIKPSVQORDPIKGFLEHVDLQVKKGETSVS EIPVQAGELAPGGLEFLVDLPVEAEATHIPQOVTVSAGLAGEAGASIHAKDIALP SGVKLDVGDGTVTLQVLSAQAEAPAEAGEGEGAAEA"

2368..2625

/gene="SCE66.03"

/note="Pfam match to entry PF01386 Ribosomal L25p, Ribosomal L25p family, score 94.50, E-value 2.2e-24"

2533..2556

/gene="SCE66.03"

/note="PS00017 ATP/GTP-binding site motif A (p-loop)"

3047..3052

3055..3657

/gene="SCE66.04"

/note="SCO3125; pth"

3055..3657

/gene="SCE66.04"

/note="SCE66.04, pth, peptidyl-tRNA hydrolase (EC 3.1.1.29), len: 200aa, strongly similar to many eg. SW:P2392 (PTH_ECOLI) peptidyl-tRNA hydrolase from

Escherichia coli (194 aa) fasta scores, opt: 449, z-score: 529.8, E(): 4.9e-22, 41.1% identity in 185 aa overlap. Contains Pfam match to entry PF01195 Pept. CRNA hydro, Peptidyl-CRNA hydrolase and Prosite match to P501196 Peptidyl-CRNA hydrolase signature 2.

```
/codon_start=1
/transl_table=11
/product="peptidyl-CRNA hydrolase"
/protein_id="CAB95918.1"
/db_xref="GI:8894758"
/db_xref="SPTREMBL:O9K3T8"
/translation="MDVTTDAGAPNLVAGNGPEPVASNRHVGVAVLLBERICA
RFRHGKAQAVVEGRIGPPGAPRRVITLAKPMSFMVNSGQVTLADFKVPGVNTV
AVHDELIDYGVRLKLGSGDNGHNGKSTIKSLGADYHVRGIGRPPGMPVADFV
LRDPSSTERKELEDFVDRADAVALVEGLIERAQSAYNS"
3085.3651
/misc_feature
/gene="SCB66.04"
/notes="Pfam match to entry PF01195 Pept. CRNA_hydro.
```

Query Match 3.0%; Score 45; DB 1; Length 28527;
Best Local Similarity 48.3%; Pred. No. 14;
Matches 126; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

```
QY 824 CATTCCGCACTTCCAGCAGAGAGTCCCGCAATTCAACGCTACCGCTCCGCAACGGCA 883
Db 11474 CATTACCGCACTCTCCCGCGGATCCCGGACCTCCCGCGGACCGCCGCACTCCGCG 11415
QY 884 TACCCGGGACAGCGCGGGGAAACCTCACCGACGAGAGGTCGACAGTCTCTCCGCGCGCG 943
Db 11414 ACCTCGGCTGGCGCCCGGCAACGTCACCTGCTGCTCGGACCGCTGCGCCACCGCC 11355
QY 944 GCATGTTGCGCGCTTGAAGAGCGCGCGCGGATGACCTCGCGCGCGGACGCGGACGAC 1003
Db 11354 GCATCACCGGCTAGACAACTGCCCGGATGCTGGAACGCGCCGCGACGTACGCGCGCC 11295
QY 1004 CAGCGCTCGCGCGGACCGGAGAGAAACAATTTGATTACTCAGACACGCTGACGG 1063
Db 11294 CCACCTCGCGCGCGGCGGACCTTCGCTGCGCCCGCGGACCGCGCTCGACCCCGC 11235
QY 1064 ACGAGACGCGCGTCCCTTCG 1084
Db 11234 ACGAGCCCTGCGACCTCTCG 11214
```

RESULT 12
SC3C9
LOCUS SC3C9 20721 bp DNA linear BCT 12-MAY-2002
DEFINITION Streptomyces coelicolor cosmid 3C9.
ACCESSION AL592262 AL645882
VERSION AL592262.2 GI:20520790
KEYWORDS AAA-like protein; dioxigenase; DNA-binding protein;
DNA-methyltransferase; glutamate synthase small subunit; hydrolase;
integral membrane protein; putative metal-transport protein;
secreted hydrolase; sugar kinase; vuvb, excinuclease ABC subunit B.
Streptomyces coelicolor A3(2).
Streptomyces coelicolor A3(2)
SOURCE ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 20721)
AUTHORS Redenbach,M., Kleser,H.M., Denaparte,D., Richner,A., Cullum,J.,
Kinashi,H. and Hopwood,D.A.
TITLE A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)
MEDLINE 97000351
PUBMED 8843436
REFERENCE 2 (bases 1 to 20721)
AUTHORS Warren,T. and Harris,D.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 20721)
AUTHORS Cerdeno,A.M., Parkhill,J., Barrell,J.B.G. and Rajandream,M.A.
TITLE Direct Submission
JOURNAL Submitted (25-JUN-2001) Streptomyces coelicolor sequencing project.

COMMENT

Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK
On May 9, 2002 this sequence version replaced gi:14571765.

Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC and Beowulf Genomics
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are
numbered using the following system eg. SC787.01c. SC (S.
coelicolor), 787 (cosmid name), .01 (first CDS), c (complementary
strand).

The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous. The length
in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon. Gene
prediction is based on positional base preference in codons using a
specially developed hidden Markov Model (Krogh et al., Nucleic
Acids Research, 22(12):4768-4778(1994)) and the FramePlot program
of Bibb et al., Gene 30:157-66(1984) as implemented at
<http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the
correct initiation codon. Where possible we choose an initiation
codon (atg, ggg, tgg or (acc) which is preceded by an upstream
ribosome binding site sequence (optimally 5-13bp before the
initiation codon). If this cannot be identified we choose the most
upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions. Cosmid 3C9 overlaps
cosmid STC54 and cosmid ST7H2.
Location/Qualifiers
1..20721

FEATURES

source

misc_feature

misc_feature

misc_feature

gene

CDS

```
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid 3C9"
/complement(248..508)
/notes="Pfam match to entry PF002710 DEAD, DEAD/DEAH box
conserved C-terminal domain, score 75.00, E-value 1.5e-18"
complement(1460..1873)
/notes="Pfam match to entry PF002710 DEAD, DEAD/DEAH box
helicase, score 27.00, E-value 1.2e-07"
complement(1742..1765)
/notes="P500017 ATP/GTP-binding site motif A (P-loop)"
complement(1946..2824)
/gene="SC3C9.02c"
/notes="SC3C9.02c"
complement(1946..2824)
/gene="SC3C9.02c"
/notes="SC3C9.02c"
complement(1946..2824)
/notes="SC3C9.02c"
possible integral membrane protein. len:
292 aa; similar to TR:O9K214 (EMBL:AL353863) Streptomyces
coelicolor putative integral membrane protein SCG8A.15c,
244 aa; fasta scores: Opt: 703 Z-score: 726.3 bits: 142.1
E(): 7.6e-33; 48.430% identity in 223 aa overlap. Contains
possible hydrophobic membrane spanning regions"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAC42741.1"
/db_xref="GI:14571767"
/db_xref="SPTREMBL:O93J14"
/translation="MNSTVDGFRYGLVTPVAVILMAGLGLGALGLRCAAVVSRSR
PGWLALGSAIGSGIWMHVFVAMKGFTEHTPIRYDLMTFASLAIVAVNGGIFLV
GYGARGTALTGTGTTGLGISMHYLGAMGMLDQGLTYNTFTVAASVVIAMVAFLA
ALMAAGVRFILWISVGAALIMGLAVGHTGMAAAVEVHLSTAEISVGSGLLELLAP
MLIGPLAFLLIAGVVMFDPILVMGRPARVAPRKRPGIPAHTEVPRTIRHPAHRPRR
```


FEATURES		Location/Qualifiers
source		1. .11738 /organism="Methanopyrus kandleri AV19" /strain="AV19" /db_xref="taxon:190192"
gene		122. .1183 /gene="MK1628"
CDS		122. .1183 /gene="MK1628" /codon_start=1 /transl_table=11 /product="Uncharacterized protein" /protein_id="AA02841.1" /db_xref="GI:1988343"
gene		/translation="MAERIVREVLREARRGVSVSDLSFSGEWMVNEITDRISRLVGS VGAPEDEVRSELTSETSKLSKPTTYLSIAETYSQVLPMGEEDADALPHDPSSEKTL DHLBGEAEFVCLASPVVSSPKDLVEGVNSLGLSRDLIELYLVAAGNEEVLLE WASLGEVREABEGREGDGLHCKVYANERLALGASWNLTVSSLSLRLARAEVHTNP KADDCVCNANYQLVSEFNSQSEAKRFFREDGLGPA1FEYRMGDPRTEVITRP ENGERMFTVLEEDNHGEFVNEREGLPVRGRVVRVHFFDLGRSRRQDRLSLRTQA REDPSGASGCRRAGREMER"
		1167. .1460 /gene="MK1629"
CDS		1167. .1460 /gene="MK1629" /codon_start=1 /transl_table=11 /product="Uncharacterized protein" /protein_id="AA02842.1" /db_xref="GI:1988344"
gene		/translation="MRWKDAAKKYKHQLTGRPYDGLVWGLDPPRSDPYVAGKLL VIRCEGSGSITVEFWDADPDWVSGSVLDGMLFDRSSCEGPTLFLVYD"
		complement (1485. .2831) /gene="MK1630"
CDS		complement (1485. .2831) /gene="MK1630" /codon_start=1 /transl_table=11 /product="tRNA cytosine-C5-methylase" /protein_id="AA02843.1" /db_xref="GI:1988345"
gene		/translation="MENQLAVKALVLETRGMPIPAVERTCRDRPIDVRSVQAFV YETIKRNLDELVAAGSDAHPEDRSPYVQQLRVGTLEMKIWRNPPAVTDCMVA1 AKRLVSGKAGFVNAVLRGVERVSKVDLEDRPWTERLALKYGEWFLVYLDLFFG DRSRVELLRANNRPPOYLIRNLKLDPLVAGVGLREYGVITEPTFLREYALVAG RGKGSRAWRBGLPDVQDLASASAAISAERGELVLYCAAPSGKTTHTERMLDEE VMAVDRSEWGLRVLERCRRLGTCVTTICRDAGLTVDDLPDVDR1LVDPPCSTTG VMNRNPDGRWKPKLERFAROMEILRLPALRIAEHCCTLVSTCSVSENEAIVR ALGEFDVKLYDAGVLSGJIEEFGREFSGYKKVRYWPFPHDTAGFFVAKMKGK"
		2858. .3598 /gene="argB" /note="MK1631"
CDS		2858. .3598 /gene="argB" /codon_start=1 /transl_table=11 /product="Acetylglutamate kinase" /protein_id="AA02844.1" /db_xref="GI:1988346"
gene		/translation="MEVIVKGSEVLDRVEDLARVIDSILVHGSGPEVSDVMERMGLE PRFVRGRTVDRETLQVMMVLAGLVNKRPLVASEGINALGSGVDGGLIAEKRS EVDGSEVDIGYGDVAKVAELLESLDAGYVYVVA PLGAGEDGTYYNNADPAA IAGAVRADRLVYLTDPGVULEDLPDETLIERVPRDEVLEEKGIYTGKVPFLZAA KMAVEACREAVITNLEGLLEGRTIVR"
		3532. .3846 /gene="MK1632"
CDS		3532. .3846 /gene="MK1632" /codon_start=1 /transl_table=11 /product="Uncharacterized protein conserved in archaea" /protein_id="AA02845.1" /db_xref="GI:1988347"
gene		/translation="MPGRDHEPGRVARGEGNHRREVAHLGEXIEKYISKLPVNLNEC KPRDRAGEBRLDLASRYFYDARYFLDRGWEAFTCLSYAMALLRAGAEEGVLDVED EV"
		3933. .4556 /gene="MK1633"
CDS		3933. .4556 /gene="MK1633" /codon_start=1 /transl_table=11 /product="Uncharacterized protein" /protein_id="AA02846.1" /db_xref="GI:1988348"
gene		/translation="MYRPVSGHGGSVEVLHPSDLPQGLTSLCKIKINGARVISRR SNIVARGRPSLEPLNDLDFGMSIAARESEVWVGDPLYRLVAREGERIVVNTDRL ACDFEETELKDEITVVEGVGDIVLACSSPVVASLKLGGDTAVHQAQVSGISLQWNI HTENNVDPLRTQCGHVMISITKKPATEKGASTEXTTAPRKVSSNI"
		complement (4615. .4690) /product="tRNA-Gly" /note="anticodon:GCC"
CDS		complement (4725. .5360) /gene="MK1634"
gene		complement (4725. .5360) /gene="MK1634" /codon_start=1 /transl_table=11 /product="zn-dependent hydrolase" /protein_id="AA02847.1" /db_xref="GI:1988349"
		/translation="MKRFVRGKGKITPSPDCNCLYLAAGDEGLILDGASGEVYDRLP GANVDVYALLTSHFPAAGDPALAGLEVGVRAEVLREGBDRLSAAYLFGRM PAYEPTFTFDEGTFPVGGEVEVLYTPGHSPGSCFLLGLDLAFTGDTVGFQPGRM LPGDNRKTCESLEBRLSTGVR8IFPGHGEYVIGEAVPALAEALRAEKD1"
gene		5403. .6257 /gene="MK1635"
		5403. .6257 /gene="MK1635" /codon_start=1 /transl_table=11 /product="N6-adenine-specific DNA methylase" /protein_id="AA02848.1" /db_xref="GI:1988350"
gene		/translation="MRPEELALAVBEALRRVFPDDPAWTHVTVSDALAHAILTNQ RKVVEGSGHNRQYLGVEAGYLAAGHDVSVSVVGAALRDETEFRYLPLKRALK EYIEEGSPGHRCEBRLGIYVVLGAREITGPNALVEYVVEGLEGAADVVPFGTGG ALAAVARGFQYVAIVRVHRAVREBLESGGVYIEADRPVDVLREREPTDILTADP YASTLREKLTSEPRPVDVAVVCHGSSWTRAVR1RGLLELFEDVEVPSKRGHEL SVCRRLRD"
		complement (6596. .7210) /gene="MK1636"
CDS		complement (6596. .7210) /gene="MK1636" /codon_start=1 /transl_table=11 /product="Uncharacterized protein specific for M.kandleri, Mk-1 family"
gene		/protein_id="AA02849.1" /db_xref="GI:1988351"
		/translation="MYRSHAPIDYNVERIVRALRVLRHRRKKOSLEVTYRTVGS TCGPVYVARRDRPRKHGRTLYLGRKENSVSFVWLVLDRREVLELARHMRNLIR SVLKTLLTEVSDLPYKKARVTLARGLAIFDARPSSPRIKDLLELTPRLESFAVAT LGMPLHVSYSLLKRVHHRKSLDEKHEVDPVGLERQMKLQRG"
gene		complement (7414. .8076) /gene="MK1637"
		complement (7414. .8076) /gene="MK1637" /codon_start=1 /transl_table=11 /product="orphan DOD family homing endonuclease" /protein_id="AA02850.1" /db_xref="GI:1988352"
gene		/translation="MTSSFGSEPGGRWLEDLRLALGLSPVRYEGDRDAVEAVAHSE GFAALLRGLQMPGKRTGRVHVPGLVRGCDRVALFLSLFGADG1ISARENRVEVG IAQSPWGSESEFLGVSLSLKRITTVYAVRLAAGSVETSVDRSRVFGLCFGRHEVE


```

misc_feature      /function="approximate span of module 3'"
97..1482
/gene="eryA"
/feature="putative"
/function="approximate span of beta-ketoacyl ACP synthase
of module 3'"
1693..1670
/gene="eryA"
/feature="putative"
/function="approximate span of acyltransferase domain of
module 3'"
3406..13921
/gene="eryA"
/feature="putative"
/function="approximate span of beta-ketoreductase domain
of module 3, possibly non-functional'"
4171..4428
/gene="eryA"
/feature="putative"
/function="approximate span of acyl carrier domain of
module 3'"
4471..10722
/gene="eryA"
/feature="putative"
/function="approximate span of acyl carrier domain of
module 3'"
4471..15847
/gene="eryA"
/feature="putative"
/function="approximate span of beta-ketoacyl ACP synthase
domain of module 4'"
6054..17026
/gene="eryA"
/feature="putative"
/function="approximate span of acyltransferase domain of
module 4'"
7165..19216
/gene="eryA"
/feature="putative"
/function="approximate span of dehydratase and
enoylreductase domains of module 4'"
9433..19984
/gene="eryA"
/feature="putative"
/function="approximate span of beta-ketoreductase domain
of module 4'"
10225..10483
/gene="eryA"
/feature="putative"
/function="approximate span of acyl carrier domain of
module 4'"
10723..20235
/gene="eryA"
/feature="putative"
/function="approximate span of acyl carrier domain of
module 4'"
6-deoxyerythronolide B formation; putative"
/codon_start=1
/transl_table=11
/protein_id="AA026495.1"
/db_xref="GI:152695"
/translation="MSGDNGTTEBKRLRYKRYTELDSVTARLREVEHRADEPIAIV
GMACRPFQDVDPSPFWEVSGGDAIAEAPADGWSLDDPARLGMILAAAGDDAF
EGISPRALAMDPOORIMLEISALBRAGDPLVLSGASVSGISGCTNHLAV
APDEVLYGVGTGSAVAGRVAYCLIGEGPAMVDPAACSGGLALHAMSLSRDC
GLAAGVTWMSRPGAFTEPRSGGLAAGCKRPFKAADGFGIABGAVLYORLSA
ARRGRPVLAIRGSANODGASNGLTAPSPACORVTRALENAGVAGVDVTEAH
GTGRIDPLEVHALSTYGERDPPDLWIGSVKSNIGHTQAAGVAGVAKAVLALR
HGEMPTLHDEBPPOIEMWDIGAVSVQASWAGRPRAAGVSSGICSTNHLAV
EAEADBPBPADSPGVPVLVSGRDOAMQAGRLADHLARPRMSLRDGTLLAT
RSAMERAVVVGVDALAGLAVAGRIADRTATGARTRGVAVWPPGCGOMOG
MAPDLRESOVFADSIDCBALAPHYDMSITDLSGARPLIDRVVOPALFAMVSL
AALRSHGVEPAVAVGHSQETLAHVAAGALITLDAKLVAVSRVLAIRLGGOGMAS
FGIETQDAERIGRFAGALSTAVNGSPVVAEGSEPLDLAECAEGITARRIPV
DYASHDPVESLREELTELIGISPVSDVAVLYSTTGOPIDITMTDIAVYANIREO
VRFQDATROLAEAGFDAFVEVSPHVLITGIEATLIDSLPADAGACVVGTLRRDGL

```

```

ADFTALGEAYAGVEVDSAPAFADRPVELPYPFQORWYLPITGCGARDEBDDM
RYQVWREAEWESASLAGRVLLVTCGVSSELSIDARSIGLEQCATVLTCDVRSSTI
GTALAEADTDALSTVVSILSRDOEAVIDPSLDALALQALGAQVEAPLWLTANAVO
ADEGILVDPQAMWGVGVGTEOPGRCMGVLDVGDADAASITSLAAVADPREGEOV
ALRADITKARLVLPAPARARTMSRGTVLVGTGCGICAHYARPLASGAHVLVL
GRRGADAPGASLEURELTALGTCVTTLAACQVARRALEVLAERAEGRVSAVMAHA
GVSTSTPLDLTAEPTETIADYVKTVMIDELCPDIDAFLVLSNAGVSGGLASV
AANAFLDGFARRRREGAPVTSIANGVLNOMQADDEGEVLRSGLRAMDPAVE
ELHITLDGOTSVDVMDRRRVELEFTHARRPLDEIYAGARAEAROSEEPALOR
LAISTAEERREHIAHLIRAEVAVLIGDDAIDDRAPADIPDSMTAVDLNRLA
VTCVREAAVTPHPPTITRLADHYLERLVGAEAEOAPALVREVPDADDPILVGM
CRPFGVHNPELMEFTVGGDAVTEMPTRDMDLALFDPDRQRTGTSRRCAGFLD
GAADPDPAEFGISPREALAMDPOORQVLLPRLAGIDPFIAGIDPHSLRSGDVGFLGAY
OGYGODAVVEDSEGYLTGNSAAVSGRAVYLGEPEVATVDTGSSSLAHASAC
GSLRDDCGLAAGVSVWAGPEVTEFGSOGGLAVDGRKATSAEDGFPAGVAV
VILORISDARRAROVLYVAGSAINODASNSGLAPSCVAAOORVIRKMAARAGITGA
DVAVVEHGTGTRLDGPVEASALLATYKSGSGSVLYLGSVSNIGHAQAAGVAVY
IKVVLGNRGLVPPMLCRGERSPLIEWSSGVLAEVASVPPAPAGVRRAGVSAGV
SGTNAAVILAEPEPEPLPEPGPVGLAANVSPVLLSARTETALAQAARLBSAVD
SVPLTALASLATGRAPHLPRRALLAGDHQULRGOLRAVEGVAACGATTTGASAGV
VFVPPGQAGMEGMARGGLSVPAFESIAECDAVLSEVAGFSASEVLEORPDAPSLER
VDVVOPLVESVMSIARLMGACGVSASVAGHOGCEIAAVVAGVLSLENGVVALR
AKALRALAGKGVNSILAEGERARALIAPEDEITSAAVNSPSSVVSQPEALAEV
ARCEDEGVRAKTLVPDYASHSRVBEIRETITLADLQISARALIPULSTLHGERDG
ADMGPRYWDNLRSOVRFDEAVSAVADGATVEVMSPHVLPAAVOELIADVAIGS
LHRDTAEHLIAELARAHVGVAVDNRNVPAAPVLPVPEPPOYMLAEVSDL
ADSRVYDMPPLATTTVDLEGFLVHGSANESITSAVERKAGVVPVPAASADRESAAL
REVPGEVAGVLSVHTGATTHLAHQSLGEGAVAPLIMLTYSRAVALGSESPVDEQAM
VWGLGRVMELETERWGLVDLPABEPAPGGEAFVACLGDHGEDOVAIRDHARYGR

```

Query Match 2.9%; Score 43.4; DB 1; Length 20235;
 Best Local Similarity 46.1%; Pred. No. 31;
 Matches 182; Conservative 0; Mismatches 211; Indels 2; Gaps 1;

```

Qy 845 AGTCCGCCGAATTCAAGCGCTACAGCTCGCAACGGGATCCCGCGAAGCCCGGAA 904
Db 12623 AGCCCGCGCGTGTGCGCCACTCGAGGCGAGATCGCCGCGACCTCGCCGCG
Qy 905 ACCTCACCCAGAGAGGTGCAAGTTCCTCGGCGCGGCGATGTTGGCGCGTTGAA 964
Db 12683 CGCTCACCTTGAGAGAGCGCCCAAGCTCGCGGCTCGGAGCGCGGCTCTGGCCG
Qy 965 GCGGCGCGCGGATTCAGCTCGCGCGAGCGGCGAGACCCAGCGCTCGGCGGCGG 1024
Db 12743 TCGGCGCGCGAGGCGGCGATGCGCTGTTGCGGCTGGGCGACCGAGCGGCGG 12802
Qy 1025 AGAGGAACAACATTTGATTACTGACACGCTGACGAGCGAGACGCGTCCCTTCG 1084
Db 12803 TCGGCGCGCTTGGCGGCGCGCTTCATCGCTCGTCAAGCGGCGCGGTCGTCG 12862
Qy 1085 GTGCACAGTGAAGAAAGCAACCTCGACAGACCTCGGCTGAGACCGGTGACACCTT 1142
Db 12863 TCGCGGGGGAAGGCGGCGGTGACGAGCTGATCGCCGAGTGGAGGCGCAAGCA 12922
Qy 1143 CCAAGCTATGCGGTCACGATATCCCGTACGGGCCAGAAACGTCTGATGAGAACTTGC 1202
Db 12923 CGGCGCGCGCGATCCCGCTCACTACGCTCCCACTACCGCAGTGAAGTCGCTGCG 12982
Qy 1203 GGGCGTGACTGCGGAAGACCGCGGCTCTTTCTTTCG 1237
Db 12983 AGGAGCTGCTGACGAGCTGCGGCGGCGATCCCG 13017

```

RESULT 15
 AR049368 20235 bp DNA linear PAT 29-SEP-1999
 LOCUS AR049368
 DEFINITION Sequence 3 from patent US 5824513.
 ACCESSION AR049368
 VERSION AR049368.1 GI:6005407
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 Unclassified.


```
FT      /*tag= C
PT      /note= "ORF 3"
XX
XX      MO9313663-A.
XX
XX      22-JUL-1993.
XX
XX      17-JAN-1992; 92MO-US00427.
XX
XX      17-JAN-1992; 92MO-US00427.
XX
XX      17-JAN-1992; 92MO-US00427.
XX
XX      (ABBO ) ABBOTT LAB.
XX
XX      Donadio S, Katz L, McAlpine UB;
XX
XX      WPI; 1993-242804/30.
XX      P-PSDB; AAR44430-32.
XX
XX      Biosynthesis of specific polypeptide analogues esp. erythromycin
PT      cpds. - by introducing altered biosynthetic gene-contg. DNA into
PT      microorganisms
XX
XX      Claim 27; Fig 2; 133pp; English.
XX
XX      This sequence represents a fragment of the Saccharopolyspora erythraea
CC      genome, designated eryA. The polypeptides encoded by this region
CC      are involved in the biosynthesis of the polypeptide segment of
CC      erythromycin. eryA is organised in modules and each module takes care
CC      of one condensation step. The precise succession of elongation steps
CC      is dictated by the genetic order of the modules. This fragment may be
CC      specifically altered such that novel polypeptide molecules of desired
CC      structure are produced. Three types of alteration may be produced:
CC      those inactivating a single function in a module which does not arrest
CC      acyl chain growth; those inactivating a single function in a module
CC      which does affect chain growth; and those affecting an entire module.
CC      The mutations may be introduced by gene replacement.
XX
XX      Sequence 29879 BP; 3672 A; 10596 C; 11545 G; 4066 T; 0 other;
SQ
Query Match      2.9%; Score 43.4; DB 14; Length 29879;
Best Local Similarity 46.1%; Pred. No. 0.13;
Matches 182; Conservative 0; Mismatches 211; Indels 2; Gaps 1;
```

```
AA097686
ID      AA097686 standard; RNA; 9757 BP.
XX
XX      AA097686;
AC
XX      27-FEB-1996 (first entry)
DT
XX      Infectious rubella virus RNA.
DE
XX      Rubella; vaccine; mutant; epitope; virus; autoimmune disease;
KM      pregnancy; foetal infection; vector; plasmid; ss.
XX
XX      Rubella virus.
OS
XX
XX      Key
FH      Location/Qualifiers
FT      CDS
FT      41..6658
FT      /tag= a
FT      /product= N-terminal transcript.
FT      misc_difference 2261..2263
FT      /tag= b
FT      /transl_except= CGU encodes Ala.
FT      misc_difference 6605..6607
FT      /tag= c
FT      /transl_except= GUC encodes asparagine or aspartic acid.
FT      misc_difference 8460..8462
FT      /tag= d
FT      /transl_except= CUG encodes Proline.
FT      misc_difference 8463..8465
FT      /tag= e
FT      /transl_except= CUC encodes Cysteine.
FT      misc_difference 9075..9077
FT      /tag= f
FT      /transl_except= UGC encodes Methionine.
XX
XX      US5439814-A.
XX
XX      08-AUG-1995.
XX
XX      28-JUN-1991; 91US-0722334.
XX
XX      28-JUN-1991; 91US-0722334.
XX      19-JUL-1993; 93US-0093453.
XX
XX      (GEOR-) GEORGIA STATE RES FOUND INC.
XX
XX      Dominguez G, Frey TK, Wang C;
PI
XX
XX      WPI; 1995-283097/37.
XX      P-PSDB; AAR79048, AAR79049.
XX
XX      New DNA encoding infectious rubella virus - esp. non-pathogenic
PT      mutant virus for use in vaccines having reduced side effects
XX
XX      Claim 2; Columns 9-18; 27pp; English.
XX
XX      Non-pathogenic mutants of the DNA corresponding to this sequence are
CC      useful in vaccines (which may include epitopes from other viruses).
CC      The mutant vaccines are less likely to cause foetal infections,
CC      autoimmune disease or neurological symptoms, so can be administered
CC      safely to older or pregnant women. The mutant sequence is
CC      preferably present in a vector, especially a bacterial plasmid that
CC      allows replication of the sequence.
XX
SQ      Sequence 9757 BP; 1458 A; 3782 C; 3008 G; 1509 U; 0 other;
Query Match      2.8%; Score 42.4; DB 16; Length 9757;
Best Local Similarity 47.8%; Pred. No. 0.15;
Matches 86; Conservative 8; Mismatches 86; Indels 0; Gaps 0;
```

RESULT 3

```
XX      841 CAGAGGTCCCGCAATTCACACGCTGCGCCCAACGGATACCCCGGAACAGCGCG 900
DB      5600 CAUACGAGCCCGAGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUG 5659
```

```
QY 901 GGAAACCTCACCAGCAGGAGGTGCAGAGTTCTCGCGCGCGCATGTTTCGGCGGTTGG 960
Db 5660 GUCGACUUCACUGAGUUGCAUGAACAGACCCUCCUACUUGGGAGUUGGACUUGAG 5719
QY 961 AAGAGCGCGCGCGCGATTGACCTCGCGCCGACGCGGACGACCCAGCGCTCGGCGCGGAC 1020
Db 5720 AUUAGCGCGCGCUCUUGGCGCCUUGCGCGGAGACUACCGCGCGCUCUUGCGCGCGG 5779

RESULT 4
AAT89642
ID AAT89642 standard; cDNA; 9759 BP.
XX
AC AAT89642;
XX
DT 26-JAN-1998 (first entry)
XX
DE Infectious rubella virus cDNA clone.
XX
KW Rubella; German measles; foetal death; birth defect; vaccine;
KW immunisation; non-pathogenic mutant; genetic engineering; ss.
XX
OS Rubella virus.
XX
PH Key Location/Qualifiers
CDS 41..6388
FT /*tag= a
FT /note= "encodes p150 and p90"
CDS 6509..9700
FT /*tag= b
FT /note= "encodes capsid, envelope 1 and envelope 2
FT /note= "glycoproteins"
XX
PN US5663065-A.
XX
PD 02-SEP-1997.
XX
PF 28-JUN-1991; 91US-0722334.
XX
PR 02-JUN-1995; 95US-0459041.
PR 28-JUN-1991; 91US-0722334.
PR 19-JUL-1993; 93US-0093453.
XX
XX (UYGE-) UNIV GEORGIA STATE RES FOUND.
XX
PI Dominguez G, Frey TK, Wang C;
XX
XX WPI; 1997-447930/41.
XX
XX DNA encoding infectious rubella virus - useful for producing
XX non-pathogenic mutants for use as vaccines
XX
XX Claim 1; Column 9-18; 19pp; English.
XX
CC AAT89642 is a cDNA molecule encoding an infectious rubella virus. The
CC sequence can be modified to give non-pathogenic mutants of this clone.
CC Such mutants are useful as vaccines for protecting against rubella
CC infection. The non-pathogenic clones can be used for the vaccination
CC of pregnant and older women with decreased risk of causing foetal
CC infection, autoimmune disease or neurological symptoms.
XX
XX Sequence 9759 BP; 1458 A; 3784 C; 3007 G; 1510 T; 0 other;
SQ

Query Match 2.8%; Score 42.4; DB 18; Length 9759;
Best Local Similarity 52.2%; Pred. No. 0.15;
Matches 94; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 841 CAGAAGTCCCGAATTCAACGGGTACACGCTCGCCAAACGCGATACCCGGGAACAGCGCG 900
Db 5600 CATACGAGCGCGCGAGGTGCGATGCGTGGCGAGGCCCATACACCAACGCGCATCGAG 5659
QY 901 GGAAACCTCACCAGCAGGAGGTGCAGAGTTCTCGCGCGCGCATGTTTCGGCGGTTGG 960
Db 5660 GTCGACTTCACTGAGTTCGACATGAACACGACCCCTCGCTACTCGGACGCTCGAGCTCGAG 5719
QY 961 AAGAGCGCGCGCGCGATTGACCTCGCGCGGAGACCCAGCGCTCGGCGCGGAC 1020
```

```
Db 5660 GTCGACTTCACTGAGTTCGACATGAACACGACCCCTCGCTACTCGGACGCTCGAGCTCGAG 5719
QY 961 AAGAGCGCGCGCGCGATTGACCTCGCGCGGAGACCCAGCGCTCGGCGCGGAC 1020
Db 5720 ATTAGCGCGCGCTCTCTTGGGCGCTTCTTGGCGGAAAGACTACCGCGCGCTCCGCGCGGC 5779

RESULT 5
AAA08043
ID AAA08043 standard; cDNA; 9759 BP.
XX
AC AAA08043;
XX
DT 19-JUN-2000 (first entry)
XX
DE Rubella virus cDNA clone sequence SEQ ID NO:1.
XX
KW Rubella virus; infection; infectious; chimeric construct; anti-viral;
KW vaccine; birth defect; autoimmune disease; ss.
XX
OS Rubella virus.
XX
PN US6054573-A.
XX
PD 25-APR-2000.
XX
PF 02-SEP-1997; 97US-0999733.
XX
PR 28-JUN-1991; 91US-0722334.
PR 19-JUL-1993; 93US-0093453.
PR 02-JUN-1995; 95US-0459041.
XX
XX (UYGE-) UNIV GEORGIA STATE.
XX
PI Abernathy ES, Pougatchev K, Frey TK;
XX
XX WPI; 2000-328366/28.
XX
XX Highly infectious rubella virus clones useful for developing a rubella
XX vaccine that can be safely administered to pregnant and older women
XX without risk of birth defects and autoimmune disease -
XX
XX Disclosure; Column 9-18; 17pp; English.
XX
CC The present invention describes nucleic acid molecules (AAA08044 and
CC AAA08045) which are fragments of the f-therien rubella virus genome.
CC AAA08044 and AAA08045 are used to replace the corresponding fragments of
CC infectious rubella virus cDNA clone with low specific infectivity
CC referred to as Robo102 to create a chimeric construct with high
CC specific infectivity, Robo302. The highly infectious rubella virus
CC clones are useful as molecular biology tools for studying rubella virus
CC and can be used for developing recombinant vaccines against rubella.
CC The rubella vaccines developed can be safely administered to pregnant
CC and older women without risk of birth defects or autoimmune disease.
CC The present sequence represents a rubella virus cDNA clone used in
CC the exemplification of the present invention.
XX
XX Sequence 9759 BP; 1458 A; 3784 C; 3007 G; 1510 T; 0 other;
SQ

Query Match 2.8%; Score 42.4; DB 21; Length 9759;
Best Local Similarity 52.2%; Pred. No. 0.15;
Matches 94; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 841 CAGAAGTCCCGAATTCAACGGGTACACGCTCGCCAAACGCGATACCCGGGAACAGCGCG 900
Db 5600 CATACGAGCGCGCGAGGTGCGATGCGTGGCGAGGCCCATACACCAACGCGCATCGAG 5659
QY 901 GGAAACCTCACCAGCAGGAGGTGCAGAGTTCTCGCGCGCGCATGTTTCGGCGGTTGG 960
Db 5660 GTCGACTTCACTGAGTTCGACATGAACACGACCCCTCGCTACTCGGACGCTCGAGCTCGAG 5719
QY 961 AAGAGCGCGCGCGCGATTGACCTCGCGCGGAGACCCAGCGCTCGGCGCGGAC 1020
```

Db 5720 ATTAGCCCGCTCTTGGGCTCCCTTGGCGGAGAACTACCGCGGCTCCGCGCGGC 5779
RESULT 6
AAH52083
ID AAH52083 standard; DNA; 2433 BP.
XX
AC AAH52083;
XX
DT 04-SEP-2001 (first entry)
XX
DE Mycobacterium tuberculosis potential drug target gene SEQ ID 137.
XX
KM Drug target; growth; organism viability; characterisation; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN WO20013517-A1.
XX
PD 17-MAY-2001.
XX
PF 13-NOV-2000; 2000MO-US31152.
XX
PR 12-NOV-1999; 99US-0165086.
XX
PR 12-NOV-1999; 99US-0165124.
XX
PR 01-FEB-2000; 2000US-0179531.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Eisenberg D, Rotstein SH, Marcotte EM,
XX
XX WPI; 2001-329193/34.
XX
DR P-PSDB; AAG81232.
XX
DR
XX
PT Identifying nucleotide or polypeptide sequence for use as drug target,
PT involves providing algorithm that analyzes a functional relationship
PT between nucleotide or polypeptide sequences, and comparing the
PT sequences -
XX
PS Disclosure; Page 145-146; 207pp; English.
XX
XX This invention relates to a method for identifying a nucleotide or
XX polypeptide sequence that may be a drug target, or essential for growth
XX or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
XX represent DNA encoding proteins AAG81096 - AAG81241. Mycobacterium
XX tuberculosis proteins which are potential drug targets. The DNA and
XX protein sequences are used to illustrate the method of the invention. The
XX method involves providing an unknown nucleotide or polypeptide sequences,
XX and comparing it to a number of sequences along with at least one
XX algorithm capable of analyzing a functional relationship between
XX nucleotide and polypeptide sequences. The method is useful for
XX characterizing the function of nucleic acids and polypeptides that may be
XX useful as a target for a drug or essential for the growth or viability of
XX an organism.
XX
SQ Sequence 2433 BP; 468 A; 820 C; 769 G; 376 T; 0 other;
Query Match 2.8%; Score 41.8; DB 22; Length 2433;
Best Local Similarity 48.2%; Pred. No. 0.11;
Matches 118; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 1015 GCCGACCCGAGAGAAACAATTTCGATTACTCAGACACGCTGACGAGACGCCG 1074
Db 1891 CGCTACGCGCGGCGGAACTACATCTACGACGACTCCAGCTCGCGAGACATCTGTTC 1950
QY 1075 TGCCC 1079
Db 1951 GGCCC 1955
RESULT 7
ABL09011
ID ABL09011 standard; cDNA; 1572 BP.
XX
AC ABL09011;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 21515.
XX
KM Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001MO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
DR P-PSDB; ABB64908.
XX
DR
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX
PS Claim 1; SEQ ID NO 21515; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL1840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ffp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 1572 BP; 365 A; 465 C; 406 G; 336 T; 0 other;
Query Match 2.8%; Score 41.4; DB 23; Length 1572;
Best Local Similarity 47.2%; Pred. No. 0.12;
Matches 126; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

```
Db 755 AGATGCCACCGTCCGCTCCACCTCGTGGAGAAGTGTGCTACGAGTTGCCTGACGGCC 814
Qy 641 CAGGAGATCAGGTGTCATGAACACTTCGGGTTCTCGACGGCATCTCGACACCCCTCAG 700
Db 815 AGGTGATCACCAATTGGCAACGAGCGCTTCGCTGCCCGCAGGCTCTGTTCCAGCCCTCGT 874
Qy 701 TCACAGGCTGGGAGACACCGCTTTC 727
Db 875 TCCTGGGCATGGAGTCGTGCGGCATCC 901

RESULT 8
ABLO9010
ID ABL09010 standard; cDNA; 4186 BP.
XX
AC ABL09010;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 21512.
XX
DE Drosophila; developmental biology; cell signalling; insecticide;
KW
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
XX 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers BW;
XX
XX WPI; 2001-656860/75.
DR P-PSDB; ABB64907.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1; SEQ ID NO 21512; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (AB857737-AB872072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4186 BP; 1092 A; 968 C; 935 G; 1191 T; 0 other;

Query Match 2.8%; Score 41.4; DB 23; Length 4186;
Best Local Similarity 47.2%; Pred. No. 0.19;
Matches 126; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

Qy 461 CCAGCAACTCGGGACGACCTCAGCCCAATGGGTGGCGCCCTTTACTGTGTACCAACATCC 520
Db 2188 CCGATTACTGTATGAAGATCTCAGCGAGCGCGCTACAGCTTTCACCAACACCGCCGAGC 2247
Qy 521 ATGGTGTCTTCTGATTGGTAGCAGCAGGACGACACTTCTTGGATCAGTTCCACGGATGATA 580
```

```
Db 2248 GTGAGATCGTGGCGGACATCAAGGAGAAGTGTGCTACGTGCTCTGGACTTCGACGAGG 2307
Qy 581 TCTCTTCGACCTTTTGGTTCTCTCCATCAGTCAAGTGCAGGCGCTCAGTGGGTCTGCCGGTC 640
Db 2308 AGATGCCACCGCTGCCGCTCCACCTCGTGGAGAAGTGTGACGAGTTGCCCTGACGGCC 2367
Qy 641 CAGGAGATCAGGCTGTCATGAACACTTCGGGTTCTCGACGSCATCTCGACGCCCTCAG 700
Db 2368 AGGTGATCACCATTTGGCAAGAGCGGCTTCGCTGCCCGCAGGCTCTGTTCCAGCCCTCGT 2427
Qy 701 TCACAGGCTGGGAGACACCGCTTTC 727
Db 2428 TCCTGGGCATGGAGTCGTGCGGCATCC 2454

RESULT 9
ABK95664
ID ABK95664 standard; cDNA; 840 BP.
XX
AC ABK95664;
XX
DT 24-SEP-2002 (first entry)
XX
DE Grass cDNA for allergen p 5 mutant #13.
XX
KW Immunoglobulin E; IgE; allergen; allergy; ss; gene; hay fever;
KW rhinoconductivitis; rhinitis; asthma; systemic anaphylaxis; mutant;
KW vaccine; antiallergic; B cell epitope.
XX
OS Phleum sp.
OS Synthetic.
XX
PN WO2002040676-A2.
XX
XX 23-MAY-2002.
XX
XX 16-NOV-2001; 2001WO-DK00764.
XX
XX 16-NOV-2000; 2000DK-0001718.
PR 16-NOV-2000; 2000US-249361P.
PR 14-JUN-2001; 2001US-298170P.
XX
XX (ALKA-) ALK-ABELLO AS.
XX
XX Holm J, Ipsen H, Nedergaard Larsen J, Spangfort MD;
XX
XX WPI; 2002-508328/54.
DR P-PSDB; ABG67048.
XX
XX New recombinant mutant allergen, useful for preventing and/or treating
PT allergy, comprises multiple mutations and reduced immunoglobulin E
PT binding affinity -
XX
XX Example 8; Page 119-120; 210pp; English.
XX
CC The invention relates to a recombinant allergen (I) which is a mutant of
CC a naturally occurring allergen, where the mutant allergen has at least
CC four primary mutations, which each reduce the specific immunoglobulin E
CC (IgE) binding capability of the mutated allergen as compared to the IgE
CC binding capability of the naturally occurring allergen, where each
CC primary mutation is a substitution of one surface-exposed amino acid
CC residue with another residue, which does not occur in the same position
CC in the amino acid sequence of any known homologous protein within the
CC taxonomic species from which the naturally occurring allergen
CC originates, and each primary mutation is spaced from each other primary
CC mutation by at least 15 Angstrom, and the primary mutations are placed
CC in such a manner that at least one circular surface region with a area
CC of 800 Angstrom2 comprises no mutation. Also included are a composition
CC comprising two or more of the recombinant allergens, where the variant
CC allergen is defined by having at least one primary mutation, which is
CC absent in at least one of the other variants, and for each variant no
CC secondary mutation is present within a radius of 15 Angstrom from each
CC absent primary mutation; a DNA sequence encoding the recombinant allergen
```

CC or its derivative, partial sequence or degenerated sequence, or a
CC sequence which hybridises to it under stringent conditions, where the
CC derivative, partial sequence, degenerated sequence or hybridising
CC sequence encodes a peptide having at least one B cell epitope; an
CC expression vector comprising the DNA and a host cell comprising the
CC vector. The recombinant allergen is useful as a pharmaceutical, for
CC preparing a pharmaceutical for preventing and/or treating allergy, or in
CC a diagnostic assay for assessing relevance, safety or outcome of therapy
CC of a subject, where an IgE containing sample of the subject is mixed
CC with the recombinant allergen and assessed for the level of reactivity
CC between the IgE in the sample and the recombinant allergen. The
CC recombinant allergen or compositions are useful for generating an
CC immune response in a subject, for vaccination or treatment of a subject
CC or for the treatment, prevention or alleviation of allergic reactions
CC in a subject e.g. hay fever, rhinoconjunctivitis, rhinitis, asthma or
CC systemic anaphylaxis. The present sequence is a cDNA encoding a
CC recombinant allergen of the invention.

XX
SQ Sequence 840 BP; 164 A; 340 C; 249 G; 87 T; 0 other;

Query Match 2.8%; Score 41.2; DB 24; Length 840;
Best Local Similarity 48.3%; Pred. No. 0.1; Mismatches 123; Indels 0; Gaps 0;
Matches 115; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 847 GTCCCGAATTCAACGGCTACAGCTCGCCAAACCGATACCGCGGAACAGCGGGAAC 906
DB 385 GCCACCGTAAAGACGGCGCTCCGCATATCGCCGACACCTTCGAGTCCACGCCGTCAAG 444
QY 907 CTCACCCAGCAGAGAGTGCAGAGTTCTCGCGCGCGCATGTTCCGCGTTGAAAGAC 966
DB 445 CCGCGCGCGAGAGGCTCAAGGTATCCCCCGCGAGCGTGCATGCAAGAGGTC 504
QY 967 GCGCGCGCGATTGACTCGCGCGCGAGCGAGACCCAGCGCTCGGCGCCAGCCCGAG 1026
DB 505 GAGCGCGCTTCAAGGTGCTGCGCCACCGCGCCAGCGCGCCCAAGCAAGATTC 564
QY 1027 AGAACAACAATTTCATTACTACAGACAGCTGACGAGACGCGCTGCCCTTCG 1084
DB 565 ACCGCTTTCAGCGCGCTTCAAGAGCGCATCAAGGAGACAGGCGGCGGCTTAG 622

RESULT 10
ID ABK95652 standard; cDNA; 861 BP.
XX
XX ABK95652;
AC
XX 24-SBP-2002 (first entry)
XX
XX
DE Grass cDNA for allergen p 5 mutant #1.
XX
XX Immunoglobulin E; IgE; allergen; allergy; ss; gene; hay fever;
KW rhinoconjunctivitis; rhinitis; asthma; systemic anaphylaxis; mutant;
XX vaccine; anti-allergic; B cell epitope.
XX
OS Phleum sp.
OS Synthetic.
XX
PN WO200240676-A2.
PD 23-MAY-2002.
XX
PF 16-NOV-2001; 2001MO-DK00764.
XX
PR 16-NOV-2000; 2000DK-0001718.
PR 16-NOV-2000; 2000US-249361P.
PR 14-JUN-2001; 2001US-298170P.
XX
XX
PA (ALKA-) ALK-ABEIL0 AS.
XX
PI Holm J, Ipsen H, Nedergaard Larsen J, Spangfort MD;
XX WPI; 2002-508328/54.
DR

DR P-PSDB; ABG67036.
XX
XX New recombinant mutant allergen, useful for preventing and/or treating
PT allergy, comprises multiple mutations and reduced immunoglobulin E
PT binding affinity -
XX
XX Example 8; Page 115; 210pp; English.
PS
XX The invention relates to a recombinant allergen (I) which is a mutant of
CC a naturally occurring allergen, where the mutant allergen has at least
CC four primary mutations, which each reduce the specific immunoglobulin E
CC (IgE) binding capability of the mutated allergen as compared to the IgE
CC binding capability of the naturally occurring allergen, where each
CC primary mutation is a substitution of one surface-exposed amino acid
CC residue with another residue, which does not occur in the same position
CC in the amino acid sequence of any known homologous protein within the
CC taxonomic species from which the naturally occurring allergen
CC originates, and each primary mutation is spaced from each other primary
CC mutation by at least 15 Angstrom, and the primary mutations are placed
CC in such a manner that at least one circular surface region with a area
CC of 800 Angstrom² comprises no mutation. Also included are a composition
CC comprising two or more of the recombinant allergens, where the variant
CC allergen is defined by having at least one primary mutation, which is
CC absent in at least one of the other variants, and for each variant no
CC secondary mutation is present within a radius of 15 Angstrom from each
CC absent primary mutation; a DNA sequence encoding the recombinant allergen
CC or its derivative, partial sequence or degenerated sequence, or a
CC sequence which hybridises to it under stringent conditions, where the
CC derivative, partial sequence, degenerated sequence or hybridising
CC sequence encodes a peptide having at least one B cell epitope; an
CC expression vector comprising the DNA and a host cell comprising the
CC vector. The recombinant allergen is useful as a pharmaceutical, for
CC preparing a pharmaceutical for preventing and/or treating allergy, or in
CC a diagnostic assay for assessing relevance, safety or outcome of therapy
CC of a subject, where an IgE containing sample of the subject is mixed
CC with the recombinant allergen and assessed for the level of reactivity
CC between the IgE in the sample and the recombinant allergen. The
CC recombinant allergen or compositions are useful for generating an
CC immune response in a subject, for vaccination or treatment of a subject
CC or for the treatment, prevention or alleviation of allergic reactions
CC in a subject e.g. hay fever, rhinoconjunctivitis, rhinitis, asthma or
CC systemic anaphylaxis. The present sequence is a cDNA encoding a
CC recombinant allergen of the invention.

XX
SQ Sequence 861 BP; 172 A; 346 C; 251 G; 92 T; 0 other;

Query Match 2.8%; Score 41.2; DB 24; Length 861;
Best Local Similarity 48.3%; Pred. No. 0.1; Mismatches 123; Indels 0; Gaps 0;
Matches 115; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 847 GTCCCGAATTCAACGGCTACAGCTCGCCAAACCGATACCGCGGAACAGCGGGAAC 906
DB 385 GCCACCGTAAAGACGGCGCTCCGCATATCGCCGACACCTTCGAGTCCACGCCGTCAAG 444
QY 907 CTCACCCAGCAGAGAGTGCAGAGTTCTCGCGCGCGCATGTTCCGCGTTGAAAGAC 966
DB 445 CCGCGCGCGAGAGGCTCAAGGTATCCCCCGCGAGCGTGCATGCAAGAGGTC 504
QY 967 GCGCGCGCGATTGACTCGCGCGCGAGCGAGACCCAGCGCTCGGCGCCAGCCCGAG 1026
DB 505 GAGCGCGCTTCAAGGTGCTGCGCCACCGCGCCAGCGCGCTCAAGCAAGATTC 564
QY 1027 AGAACAACAATTTCATTACTACAGACAGCTGACGAGACGCGCTGCCCTTCG 1084
DB 565 ACCGCTTTCAGCGCGCTTCAAGAGCGCATCAAGGAGACAGGCGGCGGCTTAG 622

RESULT 11
ID ABK95653 standard; cDNA; 861 BP.
XX
XX ABK95653;
AC
XX

comprising two or more of the recombinant allergens, where the variant allergen is defined by having at least one primary mutation, which is absent in at least one of the other variants, and for each variant no secondary mutation is present within a radius of 15 Angstrom from each absent primary mutation; a DNA sequence encoding the recombinant allergen or its derivative, partial sequence or degenerated sequence, or a sequence which hybridises to it under stringent conditions, where the derivative, partial sequence, degenerated sequence or hybridising sequence encodes a peptide having at least one B cell epitope; an expression vector comprising the DNA and a host cell comprising the vector. The recombinant allergen is useful as a pharmaceutical, for preparing a pharmaceutical for preventing and/or treating allergy, or in a diagnostic assay for assessing relevance, safety or outcome of therapy of a subject, where an IGE containing sample of the subject is mixed with the recombinant allergen and assessed for the level of reactivity between the IGE in the sample and the recombinant allergen. The recombinant allergen or compositions are useful for generating an immune response in a subject, for vaccination or treatment of a subject or for the treatment, prevention or alleviation of allergic reactions in a subject e.g. hay fever, rhinoconjunctivitis, rhinitis, asthma or systemic anaphylaxis. The present sequence is a cDNA encoding a recombinant allergen of the invention.

Sequence 861 BP; 173 A; 344 C; 252 G; 92 T; 0 other;

Query Match 2.8%; Score 41.2; DB 24; Length 861; Best Local Similarity 48.3%; Pred. No. 0.1; Matches 115; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

847 GTCCCGCAATTCAACGCTACACGCTCGCCCAACGCGATACCGCGAGAACGCGGGAAC 906
385 GCCACCGCTAAGCGAGGCGCTCAGCATATCCCGCACCTCTGAGGTCACGCGCTCAAG 444
907 CTCACCGAGCGAAGGTGAGAGTCTCTCGCGCGCGCATGTTCCGCTTGAAGAGC 966
445 CCCGCGCGCGAGGAGTCAAGGTATCCCGCGCGAGCTCAGGTATGAGAAAGTTC 504
967 GGCAGCGCGGATTAAGCTTCGCGCGCGAGCGAGCAACCGAGCTCGCGCGCGAG 1026
505 GAGCGCGCTTCAAGGTGCTGCGCACCGCCCAACGCGCGCGCGCGCAAGCAAGATT 564
1027 AGGAACAACAATTTCGATTACTCAGACACGCTGACGAGCGAGCGCGCTGCGCTTCG 1084
565 ACCGTTTCGAGCGCGCTTCAAGCAGCGCATCAAGGAGAGCAGCGGCGCGCTACG 622

RESULT 13
ABK95657
ID ABK95657 standard; cDNA; 861 BP.

XX AC ABK95657;

XX XX 24-SEP-2002 (first entry)

XX XX Grass cDNA for allergen p 5 mutant #6.

XX KM Immunoglobulin E; IGE; allergen; allergy; ss; gene; hay fever;
XX KW rhinoconjunctivitis; rhinitis; asthma; systemic anaphylaxis; mutant;
XX KM vaccine; anti-allergic; B cell epitope.

XX OS Phleum sp.
XX OS Synthetic.

XX PN MO200240676-A2.

XX PD 23-MAY-2002.

XX PF 16-NOV-2001; 2001WO-DK00764.

XX PR 16-NOV-2000; 2000DK-0001718.

XX PR 16-NOV-2000; 2000US-249361P.
XX PR 14-JUN-2001; 2001US-298170P.

PA (ALKA-) ALK-ABELLO AS.
XX Holm J, Ipsen H, Nedergaard Larsen J, Spangfort MD;
PI WPI; 2002-508328/54.
XX WPI; 2002-508328/54.
DR P-PSDB; ABG67041.
XX New recombinant mutant allergen, useful for preventing and/or treating
PT allergy, comprises multiple mutations and reduced immunoglobulin E
PS binding affinity -
XX Example 8; Page 117; 210pp; English.

The invention relates to a recombinant allergen (I) which is a mutant of a naturally occurring allergen, where the mutant allergen has at least four primary mutations, which each reduce the specific immunoglobulin E (IGE) binding capability of the mutated allergen as compared to the IGE binding capability of the naturally occurring allergen, where each CC primary mutation is a substitution of one surface-exposed amino acid residue with another residue, which does not occur in the same position in the amino acid sequence of any known homologous protein within the CC taxonomic species from which the naturally occurring allergen CC originates, and each primary mutation is spaced from each other primary CC mutation by at least 15 Angstrom, and the primary mutations are placed in such a manner that at least one circular surface region with a area CC of 800 Angstrom² comprises no mutation. Also included are a composition CC comprising two or more of the recombinant allergens, where the variant CC allergen is defined by having at least one primary mutation, which is CC absent in at least one of the other variants, and for each variant no CC secondary mutation is present within a radius of 15 Angstrom from each CC absent primary mutation; a DNA sequence encoding the recombinant allergen CC or its derivative, partial sequence or degenerated sequence, or a CC sequence which hybridises to it under stringent conditions, where the CC derivative, partial sequence, degenerated sequence or hybridising CC sequence encodes a peptide having at least one B cell epitope; an CC expression vector comprising the DNA and a host cell comprising the CC vector. The recombinant allergen is useful as a pharmaceutical, for CC preparing a pharmaceutical for preventing and/or treating allergy, or in CC a diagnostic assay for assessing relevance, safety or outcome of therapy CC of a subject, where an IGE containing sample of the subject is mixed CC with the recombinant allergen and assessed for the level of reactivity CC between the IGE in the sample and the recombinant allergen. The CC recombinant allergen or compositions are useful for generating an CC immune response in a subject, for vaccination or treatment of a subject CC or for the treatment, prevention or alleviation of allergic reactions CC in a subject e.g. hay fever, rhinoconjunctivitis, rhinitis, asthma or CC systemic anaphylaxis. The present sequence is a cDNA encoding a CC recombinant allergen of the invention.

SQ Sequence 861 BP; 171 A; 346 C; 252 G; 92 T; 0 other;

Query Match 2.8%; Score 41.2; DB 24; Length 861; Best Local Similarity 48.3%; Pred. No. 0.1; Matches 115; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

847 GTCCCGCAATTCAACGCTACACGCTCGCCCAACGCGATACCGCGAGAACGCGGGAAC 906
385 GCCACCGTACAGCGGCTTCGATATCCCGCACCTCTGAGGTCACGCGCTCAAG 444
907 CTCACCGAGCGAAGGTGAGAGTCTCTCGCGCGCGCATGTTCCGCTTGAAGAGC 966
445 CCCGCGCGCGAGGAGTCAAGGTATCCCGCGCGAGCTCAGGTATGAGAAAGTTC 504
967 GGCAGCGCGGATTAAGCTTCGCGCGCGAGCGAGCAACCGAGCTCGCGCGAG 1026
505 GAGCGCGCTTCAAGGTGCTGCGCACCGCCCAACGCGCGCGCGCGCAAGCAAGTTTC 564
1027 AGGAACAACAATTTCGATTACTCAGACACGCTGACGAGCGAGCGCGCTGCGCTTCG 1084
565 ACCGTTTCGAGCGCGCTTCAAGCAGCGCATCAAGGAGAGCAGCGGCGCGCTACG 622

RESULT 14

ABK95658
ID ABK95658 standard; cDNA; 861 BP.
XX
AC ABK95658;
XX
DT 24-SEP-2002 (first entry)
XX
XX Grass cDNA for allergen p 5 mutant #7.
DE
XX Immunoglobulin E; IgE; allergen; allergy; ss; gene; hay fever;
KW rhinoconjunctivitis; rhinitis; asthma; systemic anaphylaxis; mutant;
KW vaccine; antiallergic; B cell epitope.
XX
XX Phleum sp.
OS Synthetic.
OS
XX WO200240676-A2.
XX
PN 23-MAY-2002.
XX
PD 16-NOV-2001; 2001WO-DK00764.
XX
XX 16-NOV-2000; 2000DK-0001718.
XX
PR 16-NOV-2000; 2000US-249361P.
PR
XX 14-JUN-2001; 2001US-298170P.
XX
XX (ALKA-) ALK-ABELLO AS.
XX
XX Holm J, Ipsen H, Nedergaard Larsen J, Spangfort MD;
XX
XX WPI; 2002-508328/54.
DR
XX P-PSDB; ABG67042.
XX
XX New recombinant mutant allergen, useful for preventing and/or treating
PT allergy, comprises multiple mutations and reduced immunoglobulin E
PT binding affinity -
XX
XX Example 8; Page 117; 210pp; English.
XX
XX The invention relates to a recombinant allergen (I) which is a mutant of
CC a naturally occurring allergen, where the mutant allergen has at least
CC four primary mutations, which each reduce the specific immunoglobulin E
CC (IgE) binding capability of the mutated allergen as compared to the IgE
CC binding capability of the naturally occurring allergen, where each
CC primary mutation is a substitution of one surface-exposed amino acid
CC residue with another residue, which does not occur in the same position
CC in the amino acid sequence of any known homologous protein within the
CC taxonomic species from which the naturally occurring allergen
CC originates, and each primary mutation is spaced from each other primary
CC mutation by at least 15 Angstrom , and the primary mutations are placed
CC in such a manner that at least one circular surface region with a area
CC of 800 Angstrom ^2 comprises no mutation. Also included are a composition
CC comprising two or more of the recombinant allergens, where the variant
CC allergen is defined by having at least one primary mutation, which is
CC absent in at least one of the other variants, and for each variant no
CC secondary mutation is present within a radius of 15 Angstrom from each
CC absent primary mutation; a DNA sequence encoding the recombinant allergen
CC or its derivative, partial sequence or degenerated sequence, or a
CC sequence which hybridises to it under stringent conditions, where the
CC derivative, partial sequence, degenerated sequence or hybridising
CC sequence encodes a peptide having at least one B cell epitope; an
CC expression vector comprising the DNA and a host cell comprising the
CC vector. The recombinant allergen is useful as a pharmaceutical, for
CC preparing a pharmaceutical for preventing and/or treating allergy, or in
CC a diagnostic assay for assessing relevance, safety or outcome of therapy
CC of a subject, where an IgE containing sample of the subject is mixed
CC with the recombinant allergen and assessed for the level of reactivity
CC between the IgE in the sample and the recombinant allergen. The
CC recombinant allergen or compositions are useful for generating an
CC immune response in a subject, for vaccination or treatment of a subject
CC or for the treatment, prevention or alleviation of allergic reactions
CC in a subject e.g. hay fever, rhinoconjunctivitis, rhinitis, asthma or
CC systemic anaphylaxis. The present sequence is a cDNA encoding a

CC recombinant allergen of the invention.
XX
SQ Sequence 861 BP; 170 A; 345 C; 254 G; 92 T; 0 other;
XX
XX Query Match 2.8%; Score 41.2; DB 24; Length 861;
XX Best Local Similarity 48.3%; Pred. No. 0.1;
XX Matches 115; Conservative 0; Mismatches 123; Indels 0; Gaps 0;
XX
QY 847 GTCCTCCGATTAACCGCTACACGCTCCGCAACGCGATACCCGCAACAGCGGGAAC 906
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 385 GCCACCGTAAGCAGCGCTCCGCTATCATCGCGGCGCACCTCGAGTCCACGCGTCAAG 444
QY 907 CTCACCCAGCAGAGGTGCGAGTTCCTCGGCGCGCATGTTCCGCGCTTGGAGAGC 966
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 445 CCGCGCGCAGAGGTCAAGGTCTATCCCGCGCGGAGCTGCAGGTCTATCGAAGAGGTC 504
QY 967 GCGCGCGCTGACCTCGCGCGCGGCGGACGACCCAGCGCTCGGCGCGCACCGCGAG 1026
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 505 GAGCGCGCTTCAAGGTGCTGCGCACCGCGGCGGCGCGCGCGCGCAACAGATTT 564
QY 1027 AGGAACAACAAATTCGATTACTCAGACGCTGACGAGCAGACGCGCTGCCCTTCG 1084
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 565 ACGTCTTCGAGCGCGCTTCAACGACGCGCATCAAGCGCAGCAGCGCGCGCTACG 622
XX
XX RESULT 15
XX ID ABK95659
XX ID ABK95659 standard; cDNA; 861 BP.
XX
XX AC ABK95659;
XX
XX DT 24-SEP-2002 (first entry)
XX
XX DE Grass cDNA for allergen p 5 mutant #8.
XX
XX Immunoglobulin E; IgE; allergen; allergy; ss; gene; hay fever;
KW rhinoconjunctivitis; rhinitis; asthma; systemic anaphylaxis; mutant;
KW vaccine; antiallergic; B cell epitope.
XX
XX OS Phleum sp.
OS Synthetic.
XX
XX WO200240676-A2.
XX
XX 23-MAY-2002.
XX
XX 16-NOV-2001; 2001WO-DK00764.
XX
XX 16-NOV-2000; 2000DK-0001718.
XX
XX 16-NOV-2000; 2000US-249361P.
XX
XX 14-JUN-2001; 2001US-298170P.
XX
XX (ALKA-) ALK-ABELLO AS.
XX
XX Holm J, Ipsen H, Nedergaard Larsen J, Spangfort MD;
XX
XX WPI; 2002-508328/54.
DR
XX P-PSDB; ABG67043.
XX
XX New recombinant mutant allergen, useful for preventing and/or treating
PT allergy, comprises multiple mutations and reduced immunoglobulin E
PT binding affinity -
XX
XX Example 8; Page 118; 210pp; English.
XX
XX The invention relates to a recombinant allergen (I) which is a mutant of
CC a naturally occurring allergen, where the mutant allergen has at least
CC four primary mutations, which each reduce the specific immunoglobulin E
CC (IgE) binding capability of the mutated allergen as compared to the IgE
CC binding capability of the naturally occurring allergen, where each
CC primary mutation is a substitution of one surface-exposed amino acid
CC residue with another residue, which does not occur in the same position
CC in the amino acid sequence of any known homologous protein within the
CC taxonomic species from which the naturally occurring allergen
CC originates, and each primary mutation is spaced from each other primary
CC mutation by at least 15 Angstrom , and the primary mutations are placed
CC in such a manner that at least one circular surface region with a area
CC of 800 Angstrom ^2 comprises no mutation. Also included are a composition
CC comprising two or more of the recombinant allergens, where the variant
CC allergen is defined by having at least one primary mutation, which is
CC absent in at least one of the other variants, and for each variant no
CC secondary mutation is present within a radius of 15 Angstrom from each
CC absent primary mutation; a DNA sequence encoding the recombinant allergen
CC or its derivative, partial sequence or degenerated sequence, or a
CC sequence which hybridises to it under stringent conditions, where the
CC derivative, partial sequence, degenerated sequence or hybridising
CC sequence encodes a peptide having at least one B cell epitope; an
CC expression vector comprising the DNA and a host cell comprising the
CC vector. The recombinant allergen is useful as a pharmaceutical, for
CC preparing a pharmaceutical for preventing and/or treating allergy, or in
CC a diagnostic assay for assessing relevance, safety or outcome of therapy
CC of a subject, where an IgE containing sample of the subject is mixed
CC with the recombinant allergen and assessed for the level of reactivity
CC between the IgE in the sample and the recombinant allergen. The
CC recombinant allergen or compositions are useful for generating an
CC immune response in a subject, for vaccination or treatment of a subject
CC or for the treatment, prevention or alleviation of allergic reactions
CC in a subject e.g. hay fever, rhinoconjunctivitis, rhinitis, asthma or
CC systemic anaphylaxis. The present sequence is a cDNA encoding a

CC taxonomic species from which the naturally occurring allergen
CC originates, and each primary mutation is spaced from each other primary
CC mutation by at least 15 Angstrom , and the primary mutations are placed
CC in such a manner that at least one circular surface region with a area
CC of 800 Angstrom^2 comprises no mutation. Also included are a composition
CC comprising two or more of the recombinant allergens, where the variant
CC allergen is defined by having at least one primary mutation, which is
CC absent in at least one of the other variants, and for each variant no
CC secondary mutation is present within a radius of 15 Angstrom from each
CC absent primary mutation; a DNA sequence encoding the recombinant allergen
CC or its derivative, partial sequence or degenerated sequence, or a
CC sequence which hybridises to it under stringent conditions, where the
CC derivative, partial sequence, degenerated sequence or hybridising
CC sequence encodes a peptide having at least one B cell epitope; an
CC expression vector comprising the DNA and a host cell comprising the
CC vector. The recombinant allergen is useful as a pharmaceutical, for
CC preparing a pharmaceutical for preventing and/or treating allergy, or in
CC a diagnostic assay for assessing relevance, safety or outcome of therapy
CC of a subject, where an IGE containing sample of the subject is mixed
CC with the recombinant allergen and assessed for the level of reactivity
CC between the IGE in the sample and the recombinant allergen. The
CC recombinant allergen or compositions are useful for generating an
CC immune response in a subject, for vaccination or treatment of a subject
CC or for the treatment, prevention or alleviation of allergic reactions
CC in a subject e.g. hay fever, rhinoconjunctivitis, rhinitis, asthma or
CC systemic anaphylaxis. The present sequence is a cDNA encoding a
CC recombinant allergen of the invention.
XX
SQ Sequence 861 BP, 170 A, 348 C, 251 G, 92 T, 0 other;

Query Match 2.8%; Score 41.2; DB 24; Length 861;

Best Local Similarity 48.3%; Pred. No. 0.1;

Matches 115; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 847 GTCCCGGATTCAGAGCGCTACACGCTCGCCAGCGGATACCGCGAGACAGCGCGGAAC 906
DB 385 GCCACCGTAACAGCGCGCTCGCATCATCGCGGACCCCTGAGGTCCACGCCCTCAAG 444
QY 907 CTCACCCAGCAGAGAGTGCAGAGTCTCTGGCGCGCGCATGTTCCGCGCTTGAAGAGC 966
DB 445 CCGCGGCGGAGAGGTCAAGGTATCTCCCGCGGAGCTGCAAGTATCAGAGAGGTC 504
QY 967 GCGCGCGCGATTCAGCTCGCGCGGAGCAGACCCAGCGCTCGCGCGCGAGCCCGCAG 1026
DB 505 GACGCGCGCTTCAGAGTGTGCGCCACCGCCGCAAGCGCGCGCCCAAGCAAGATT 564
QY 1027 AGGAACAACAATTGATTACTCAGACACGCTGACGAGCAGAGCGCGCTGCGCTTCG 1084
DB 565 ACCGTCTTCGAGCGCGCTTCACGACGCCATCAAGGAGCAGCGCGCGCGCTTACG 622

Search completed: December 4, 2002, 21:53:19
Job time : 334 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2002, 14:49:27 ; Search time 1941 Seconds
(without alignments)
12465.769 Million cell updates/sec

Title: US-09-926-084-8

Perfect score: 1494
Sequence: 1 acgcgccttcgcctgttgc.....tcactgcgactatcgcctgc 1494

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pla:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	86.6	5.8	785	12	BG365697
2	83.2	5.6	562	9	AJ407445
3	51.4	3.4	844	17	CNS0052P
4	49.8	3.3	925	17	CNS0091P
5	46	3.1	925	17	CNS0091P
6	45	3.0	709	12	BE718112

7	44.6	3.0	677	12	BE718110
C	43.4	2.9	1101	17	CNS0175Y
8	43	2.9	608	12	BE718106
9	43	2.9	651	12	BE718108
10	43	2.9	1101	17	CNS0175Y
11	42.6	2.9	640	12	BE718105
12	42.6	2.9	760	10	AM600680
13	42	2.8	1201	17	CNS016BR
14	41.8	2.8	615	13	BU309091
15	41.4	2.8	620	12	BE718104
16	41.4	2.8	662	12	BE718109
17	41.4	2.8	504	12	PE577L
18	41.2	2.8	747	9	AJ446237
19	41.2	2.8	635	10	BE519935
20	41	2.7	976	17	CNS010DV
21	41	2.7	1201	17	CNS016BR
22	41	2.7	682	9	AJ448030
23	40.6	2.7	688	9	AJ448030
24	40.6	2.7	688	9	AJ448030
25	40.6	2.7	690	9	AJ446034
26	40.6	2.7	732	9	AJ447834
27	40.4	2.7	517	13	BM289588
28	40.4	2.7	544	13	BM289588
29	40.2	2.7	623	14	BO801635
30	40.2	2.7	932	17	CNS0072Q
31	40.2	2.7	1100	17	CNS016KD
32	39.8	2.7	625	13	BM628689
33	39.8	2.7	637	12	BE718111
34	39.8	2.7	705	13	BM652600
35	39.6	2.7	386	9	AJ106185
36	39.6	2.7	779	10	BE387585
37	39.4	2.6	839	17	CNS004NB
38	39.2	2.6	910	17	CNS006ON
39	39	2.6	478	10	BB858106
40	39	2.6	515	9	AJ354183
41	39	2.6	2229	11	AY105265
42	38.8	2.6	1201	17	CNS014BJ
43	38.6	2.6	528	9	AA464358
44	38.6	2.6	843	17	AG157675
45	38.4	2.6	568	14	BU038401

ALIGNMENTS

RESULT 1
LOCUS BG365697 785 bp mRNA EST 22-OCT-2001
DEFINITION HVSMF10003M12f Hordeum vulgare 20 DAP spike EST library HVCDNA00310
(20 DAP) Hordeum vulgare cDNA clone HVSMF10003M12f, mRNA sequence.
ACCESSION BG365697
VERSION BG365697.1 GI:113254796
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticeae; Hordeum.
1 (bases 1 to 785)
Wing,R., Close,T.J., Kleinbols,A., Wise,R., Begum,D., Frisch,D., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton,R.D., Close,S.J., Oates,R. and Main,D.
Development of a genetically and physically anchored EST resource for barley genomics: Morex 20 DAP spike cDNA library
Unpublished (2001)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Total hg bases = 490
Seq primer: AATTAACTCTACTAAAGG

```
High quality sequence stop: 543.
Location/Qualifiers
1. 785
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEI0003M12f"
/clone_lib="Hordeum vulgare 20 DAP spike EST library
HVCDA0010 (20 DAP)"
/tissue_type="20 DAP spike"
/lab_host="SOLR"
/notes="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;
plants were grown in the greenhouse at the University of
California, Riverside (Penton, SJ Close, TJ Close). Whole
spikes with awns trimmed were collected at 20 DAP (Penton
et al.). Total RNA was prepared, poly(A) RNA was purified, one
primary unamplified cDNA library was made, and 1 million
pfu were in vivo excised to give pBluescript SK(-) cDNA
phagemids in the TJ Close lab at the University of
California, Riverside (Choi). Phagemids were plated and
picked at the Clemson University Genomics Institute (CUGI)
(Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA
preparations, DNA sequencing and sequence analysis were
performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates
, Rambo, Main). The sequence has been trimmed to remove
vector sequence and contains a minimum of 100 bases of
phred value 20 or above. For more details on library
preparation and sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders Also
see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT      206 a      231 c      207 g      141 t
ORIGIN
Query Match      5.8%; Score 86.6; DB 12; Length 785;
Best Local Similarity 53.0%; Pred. No. 9.e-13;
Matches 342; Conservative 0; Mismatches 269; Indels 34; Gaps 6;

QY 633 TGCGGTCGAGGAGATCAGGCTGCTATGAACACTTCGGGTTCCTCCAGCGCATCTCGCA 692
Db 5 TGTCGCCGCCGGAATTCACAAGGGCAAGGACCTTCGGATTCAAGATGGATCACTCA 64

QY 693 GCCCTCAGTCACAGGCTGGGAGACGACCGCTTCTCCCTGGACAGCGGCTCGTCCCA 747
Db 65 GCCTCTTTTGGGGCTGGGATGAGAAGGTGCCCGGCAAGGAGCTGTGGCCACAAA 124

QY 748 -CCTGGAATATCTCACTGACCGCATGGGACACGGGACCGGACCGGCTGTGGGCTCT 806
Db 125 GCCAGGAATATCGTTTTCGGGATACGACGGCGAGCATGCAACAGCCCGCATGGGCCAA 184

QY 807 AGATGGAGTTTATGTCATTCGGCATTCAGCAGCAAGGTCCTCCGAATTCACGGTA 866
Db 185 GGACGGCAGTTTCTGCTTTCGGTGATCTGCAACAGCAGGTTCTTGATTCAGGAATT 244

QY 867 CAGCTCGCAAGCGCATATCCCGGAAACAGCGCGGGAACCTTCACCCAGCAGGAAGTGC 926
Db 245 CATGAACCAAAACGCTAAGCAGCAGCATTCACCAAGGACCAACCCCAAGC-----CAGG 298

QY 927 AGAGTTCTCTGGCGCGCATATGTTGGCGGTTTGAAGAGCGGCGCGCATGACCTCGC 986
Db 299 CGAGAAGCTCGCCGCTTACCTGATGGGTGGTGAAGAACGGAACCCCGCTTACAGGATC 358

QY 987 GCCAGCGGAGACCCAGGCTCGGCGCGACCCCGCAGAGGAACAAATTCGATTA 1046
Db 359 TCCTCAGGATGACACCAACAGAGGCTTTGTTGCAATC-----CAACAACTTCGACTT 409

QY 1047 CTGACAGACGCTACGAGCAGAGCGGCTGCCCTTCGGTGACACGTCGAGGAAGCAA 1106
Db 410 CGCA---CCGGTCACCGAGCAACAAGTGTCCCTTCAGCGGCACACATCCGCAAGATGAG 466
```

```
QY 1107 CCCTCGACAGGACCTGGGTGGACCGGTCGACACCTTTCACGCTATCGGTCCAGTATCCC 1166
Db 467 ACCCGTGTGACTTGG-----AACAGCACACGCGGTTCATCATCGTCCGGAATCCC 520

QY 1167 GTACGCCCCAGAAACGCTCTGATCGAAGACTTGGTCGGCGGTGACTGCCAA-----GACC 1222
Db 521 CTACGGCGAGTGCTGCTGACCTGAACTGGCGCGCAGAAAAACCCAAAACAAGGAGC 580

QY 1223 GCGGCTCTCTTTTTCGTCGAGTACCAGTCCATTTATTTGTTGTAATGGGT 1267
Db 581 GTGGCCTCTCTTCTGTTTGTCTACCAGATCGATTTTCGAAAGGAT 625

RESULT 2
LOCUS      AJ407445          552 bp      mRNA      linear      EST 08-JAN-2002
DEFINITION      AJ407445 Populus tremula x P. tremuloides/Amanita muscaria mixed
                  EST library Populus tremula x P. tremuloides/Amanita muscaria mixed
                  EST library cDNA clone 441, mRNA sequence.
ACCESSION      AJ407445
VERSION        AJ407445.1  GI:18091330
KEYWORDS        EST.
SOURCE          Populus tremula x P. tremuloides/Amanita muscaria mixed EST
                  library.
ORGANISM        Populus tremula x P. tremuloides/Amanita muscaria mixed EST library
                  Eukaryota; mixed EST libraries.
REFERENCE      1 (bases 1 to 552)
AUTHORS        Nehls,U., Bock,A. and Hampf,Rd.
TITLE          ESTs from Amanita muscaria / Populus tremula x tremuloides
                ectomycorrhizas
JOURNAL        Unpublished (2001)
CONTACT        Nehls U
COMMENT        Botanisches Institut/Physiologische Oekologie der Pflanzen
                  Universitaet Tuebingen
                  Auf der Morgenstelle 1, Tuebingen 72076, Germany.

FEATURES
            Location/Qualifiers
            source
            1..552
               /organism="Populus tremula x P. tremuloides/Amanita
               muscaria mixed EST library"
               /db_xref="taxon:143775"
               /clone="441"
               /clone_lib="Populus tremula x P. tremuloides/Amanita
               muscaria mixed EST library"
BASE COUNT      124 a      158 c      135 g      132 t      3 others
ORIGIN
```

```
Query Match      5.6%; Score 83.2; DB 9; Length 552;
Best Local Similarity 56.2%; Pred. No. 7.e-12;
Matches 239; Conservative 0; Mismatches 174; Indels 12; Gaps 4;

QY 623 TCAGTGGGTCTGCGCTCCAGGAGATCAGGCTGGTGCATGAACACTTCGGGTTCCTCGACG 682
Db 101 TTACTGGAGATGTAAAGCCAGGAGGCTGGACATGAACACTTCGGCTTCTTGNAC 160

QY 683 GCA--TCTCGACGCTCAGTCACAGCTGGGAGACGCGTTCCTCCGAGAGCGGT 740
Db 161 GGAATCTTCAACCGCGCGCTCATCGGATTCGATACGAATTCCTCCCTCTGAGCTCTCC 220

QY 741 CGTCCACCTTGAATTTATCTCTACTCGAGCGATGGGACACGG---GACCCGACCGTC 797
Db 221 TGTACGCTCTGGGATCATCTCTCGGCATGATGGGAGCATGTATCTACTCGCATCC 280

QY 798 GTGGGTCTAGATGGAGTTTCATGGCATTCGCGCATTCCTCCAGCAAGAGTCCCGAAT 857
Db 281 CTGGATGGTAGACGGAAGTTCTCTGCTTCCGTTACCTGTTCCAGTCGTTCCCGAAT 340

QY 858 CAACGGTACACGCTGCCAACCGGATACCCGCAACAGCGCGGGAACCTCACCAGCA 917
Db 341 CAA--CAAAGTTTCTCTCCAGGACAAATCAACGTGCTGTTGTGTGACAGCAG-- 396

QY 918 GGAAGGTGCAGAGTTCCTCGCGCGCGCATGTTTCGGCGGTTTGAAGAGCGGCGCGCAT 977
Db 397 ---AGTAGTGAATTCCTCGGTGCTCGACTTGTATGTCTGCTGGAAGCGGTGCTCCAGT 453
```

DB	Accession	Source	Organism	Reference	Authors	Title	Journal	Comment
Oy	978	TGACCTCGGCGCGAGCGGACGACCCAGCGCTTGCGCGCCGACCGCGAGGAACAACAA	844 bp	DNA	linear	GSS 03-JUN-1999		
Db	454	TGATTTGGCACTTTGACGACGATCAGCCCTTGCGCGCGATGCGCATGACGACATTA	513					
Oy	1038	TTTCG 1042						
Db	514	CTTTC 518						
RESULT 3								
CNS0052P								
CNS0052P								
LOCUS								
DEFINITION								
ACCESSION								
VERSION								
KEYWORDS								
SOURCE								
ORGANISM								
REFERENCE								
AUTHORS								
TITLE								
JOURNAL								
COMMENT								
FEATURES								
source								
BASE COUNT								
ORIGIN								
Query Match								
Best Local Similarity								
Matches								
Oy	763	ACTGACGCGATGAGGACACCGGACCGGACCGTCTGCGCTCTAGATGGAGTTTCAG	822					
Db	391	WCACGCGSSSSASBRMSSSCSSACSSSVSSSSCSAMSSSCCVSSCGMASSSCG	450					
Oy	823	GCATTCGGGACCTTCCACGACGAAAGTCTCCCGAATTTCACCGCTACCGCTCCGCAACGCG	882					
Db	451	GMSASSSSSSGVSVSVGGRAVGRCMCCSNVCCMCSKCCSVCVCAVCSGS	510					
Oy	883	ATACCGCGCAACAGCCCGGAAACCTCACCCAGCAGAGAGTGCAGAGTTCTCGGCGCG	942					
Db	511	VASVAVGCGVGRVGSNSGRAGRSSGRCGSSVSSGTVSSSSSSVCGACSSASVGS	570					
Oy	943	CGCATGTTGCGCGTTGGAAGACGGGCGCGCCGATTTGACCTGCGCGGACGCGGACGAC	1002					

D	b	571	BBSVASSVSGBVBNSGCCRCYCGVGCGGSRVSSCGSSSSSSSGCCSVKSRASGANG	630
Oy	1003	CCAGCGCTGGGCGCCGACCGCAGAGAACAATTCGATTACTGACACCGTGCAGC	1062	
D	631	VIVGSGRRGGGGGGRRGANABADDAABAABRABRANSNADAWCAVAVSSVSSBGAR	690	
Oy	1063	GACGAGACCGCGCTGCCCCCTTGCGTG	1087	
D	691	GRRAAAVNVGSVGRBNKKRYTKTK	715	
R	E	S	T	
L	C	N	S	
D	LOCUS	CNS0091P/c		
D	DEFINITION	CNS0091P	925 bp	DNA linear GSS 03-JUN-1999
D	DEFINITION	Drosophila melanogaster genome survey sequence TE73 end of BAC #		
D	DEFINITION	BACR19D16 of RPci-98 library from Drosophila melanogaster (fruit		
D	DEFINITION	fly), genomic survey sequence.		
A	ACCESION	AL053013		
V	VERSION	AL053013.1	GI:4934461	
K	KEYWORDS	GSS.		
S	SOURCE	Drosophila melanogaster.		
O	ORGANISM	Drosophila melanogaster.		
R	REFERENCE	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;		
A	AUTHORS	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
T	TITLE	Ephydroidea; Drosophilidae; Drosophila.		
J	JOURNAL	1 (bases 1 to 925)		
		Genoscope.		
		Direct Submission		
		Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :		
		BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cne.fr		
		- Web : www.genoscope.cns.fr)		
		Determination of this BAC-end sequence was carried out as part of a		
		collaboration with the Berkeley Drosophila Genome Project (BDGP).		
		The BDGP is constructing a physical map of the Drosophila		
		melanogaster genome using these BACs. For further information		
		please see http://www.fruitfly.org The BDGP Drosophila		
		melanogaster BAC library was prepared by Kazutoyo Osoegawa and		
		Aaron Mammossier in Pieter de Jong's laboratory in the Department of		
		Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,		
		NY. The library is named RPci-98 and was constructed by partial		
		EcoRI digestion of Drosophila DNA provided by the BDGP from the		
		isogenic strain Y2; cn bw sp, the same strain used for the BDGP's		
		P1 and EST libraries. A more detailed description of the library		
		and how to order individual BAC clones, the entire library, or		
		filters for hybridization from the BACPAC Resource Center can be		
		found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.		
F	FEATURES	Location/Qualifiers		
S	source	1..925		
		/organism="Drosophila melanogaster"		
		/db_xref="taxon:7227"		
		/clone="BACR19D16"		
		/clone_lib="RPci-98"		
		/note="end : TE73"		
B	BASE COUNT	120 a 61 g 172 t 511 others		
O	ORIGIN			
Q	Query Match	3.3%; Score 49.8; DB 17; Length 925;		
B	Best Local Similarity	15.4%; Pred. No.0.016; Mismatches 148; Indels 0; Gaps 0;		
M	Matches	54; Conservative 149; Mismatches 148; Indels 0; Gaps 0;		
Oy	687	CTTCGAGCCCTTAGTCACAGGCTGTGGAGACGACCTCTTCTCGTGACAGCGGTGCTCC	746	
D	900	SNSSSCSSSSBBSSSSSTSMSS	841	
Oy	747	ACCTGAATTATCTCATCTGACGCGGATGGGACACGGGACACCGACCGTGTGGCTCT	806	
D	840	BCCMCSSSSSCCGSAAGVKVRAGGAGKGGGGGSAshSSSSAACBSSSSSCSACW	781	
Oy	807	ACATGGAGATTTCATGAGCTTCGCGCACCTTCACAGAGAGATCCCGAATTAACAGCGTA	866	
D	780	SASSSSSSSSSRBRGGAGGGSASSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS	721	
Oy	867	CACGCTCGCAACGAGATACCCGCAACAGCGCGGGAATACTTACCCGACGGAAGTGC	926	

Oy	CCGCAAACTTCGGGGAGCAGCCTCAGCGAATGGGTGGCCGCTTTTACCTGGTAACAACATCC	520
Db	349 CCGATTACTGTAAGAATCTCTGACGGAGCGCGGCTACAGCTTCAACACACCACCGCCGAGC	408
Oy	521 ATGGTGCTCTTCTCATTTGGTAGCGAACGAGACGACTTTGGATGATTCACGATGATA	580
Db	409 GTGAGATCGTGCAGACATCAAGGAAGAAGCTGTGCTACGTGGCTCTGGACCTTGAGACAG	468
Oy	581 TCTCTTCCACCTTGGTTCCTCATTCACACAGTGCAGGCGCTCAGTGGTGTGGCGCTC	640
Db	469 AGATGCCACCGCTGCGCTCTCACCTGCTGGAAGAAGTCGATCAAGATTGCTTGACGGCC	528
Oy	641 CAGGAGATCAGGCTGATGAACAATTGGGTTCTTCGACGGCAATCTCGACGCCCTCAG	700
Db	529 AGGTGATCAACATTGGCAAAGCGCTTCCGCTGCCCGAGGCCCTGTGTCAGGCCCTCGT	588
Oy	701 TCACAGGCTGGGAGACGACCGTCTTCCCTGGACAGGC	737
Db	589 TCTCTGGCATGGAGTCGTGCGGCATCAAGAACCGTC	625

RESULT 7							
LOCUS	BE718110						
DEFINITION	RC1-H07097-210600-021-505	677 bp	mRNA	linear	EST 12-SEP-2000		
ACCESSION	BE718110						
VERSION	BE718110.1						
KEYWORDS	EST.						
SOURCE	human.						

REFERENCE AUTHORS

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.

Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Piof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/genhtml2.pl?l=et2-RC1-HT0797-210>)
600-021-g05&c3=2000-06-21&c4=1)
Seq primer: puc 18 forward
High quality sequence stop: 662.

FEATURES
SOURCE

```
/organism="Homo sapiens"  
/db_xref="taxon:9606"  
/clone_lib="HT0797"  
/dev_stage="Adult"
```

/note="Organ: head neck; Vector: puc18; Site 1: Sma1;
Site 2: Sma1; A mini-library was made by cloning products
derived from ORS7ERS PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research).

BASE COUNT
ORIGIN

Query Match	3.0%;	Score 44.6;	DB 12;	Length 677;
Best Local	Similarly	47.6%;	Pred. No. 0.36;	
Matches 131;	Conservative	0;	Mismatches 144;	Indels 0;
				Gaps 0;

Oy	461	CCGAAACCTCGGGAGCAGCACTCAGCCAAATGGGTGGCGCTTTTACTGTGATCACCATTCC	520
Db	347	CCGATTACTCTGATGAAAGATCCTGCAGGAGCGCGCTACAGCTTCAACACACACCGCCGAC	406
Oy	521	ATGTGTCCTTTCTGATTGTAAGCGACGAGCAGCTTCTTGATAGTTCACGGATGATA	580
Db	407	GTGAGATGATGTGGCGACATCAAGGAGAACTGTGCTTACGTGGCTCTTGACCTTCGAGCAG	466
Oy	581	TCTCTTCAGCCTTGTGTTCTCTCATCATCAGAGTGCAGAGCGCTGCAGTGTGAGCGCTC	640
Db	467	AGATGGCCACCGCTGCTCGCGCTTCACCTCGCTGGAAAGTCGTACAGAGTTGCTTGACGGCC	526
Oy	641	CAGGAGATCAGGCTGTGATGAAACATTGGGTTCTTGACGGCAATCTCGACGCTCTAG	700
Db	527	AGGTATATCACCATTGTGGCAACGAGCCCTTCCGTGCCCCGAGGCGCCTGTTCAGCCCTCGT	586
Oy	701	TCACAGCGCTGGAGAGCAGACCGTCTTCCCTGGACAG	735
Db	587	TCTGTGGCATGGAGTCTGTGGCATTCACAGAGCCG	621

RESULT 8

LOCUS	CNS017SY	1101 bp	DNA	linear	GSS 26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence SP6 end of BAC BACN37108 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.				

VERSION	AL108460.1	GI:5628764
KEYWORDS	GSS.	
SOURCE	Drosophila melanogaster	
ORGANISM	Drosophila melanogaster	

REFERENCE
1 (bases 1 to 1101)

JOURNAL Submitted (22-JUL-1999) Genoscope - Centre National de Séquençage
BP 191 91006 EVRY cedex - FRANCE (E-mail : segr@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a

collaboration with the European Drosophila Genome Project (EDGP) - <http://www.edgp.ebi.ac.uk> - This *Drosophila melanogaster* BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.

FEATURES
source

```

BASE COUNT      254 a      176 c      160 g      152 t      359 others
ORIGIN
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACN37L08"
/clone_1b="DrosBAC"
/plasmid="pBelosBAC11"
/note="end : SP6"

```

Query Match	2.9%	Score 43.4	DB 17	Length 1101
Best Local Similarity	15.7%	Pred. NO. 1		
Matches 70	Conservative 182	Mismatches 195	Indels 0	Gaps 0

```
Oy      579 TATCTTCGACCCITTTGGTTCCTCATCACTCAGGTGCAGGCCCTAAGTGCGTTCGCCG 638
        ||::||::|||::||::||::||::||::||::||::||::||::||::||::||::||
Db     1095 TTBSGTSTSSSTTTTCTTKTTBTBTTBSBSTSTSTSTBTBSBSSSBBSBBSSGSSSBSB 1036
```


/clone_1lb="HT0797"
 /dev_stage="Adult"
 /note="Organ: head, neck; Vector: puc18; Site 1: Smal;
 Site 2: Smal; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."
 BASE COUNT 124 a 185 c 212 g 130 t
 ORIGIN

Query Match 2.9%; Score 43; DB 12; Length 651;
 Best Local Similarity 47.6%; Pred. No. 0.99;
 Matches 127; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 461 CCGGAACTCTGGGAGCAGCTCAGCCATGAGTGGCCCTTTACTGTACCACTCC 520
 DB 324 CCGATTACCTGATGAATCTGACGAGCGCGCTACAGCTTACCAACCCCGCGAGC 265
 QY 521 ATGGTCTCTTCTGATTGTAGCGACGACGAGCACTTCTGATCAGTTACGGATGATA 580
 DB 264 GTGAGATCTGCGGACATCAAGAGAGCTGTCTAGCTGAGCTTCTGACTTTCAGCAGG 205
 QY 581 TCTCTGACCTTGTGTCTCTCATCACTCAGTGCAGCGCTCAGTGGTCTGCGCGTC 640
 DB 204 AGATGGCAGCGCTGCGCGCTTCCAGCTGCTGAGAGAGTCTAGAGATTGCTGACGCGC 145
 QY 641 CAGAGATCAGAGCTGTATCAACACTTGGGTTCTTGCAGCGATCTCGAGCCCTTACG 700
 DB 144 AGGATACCATTTGGAGAACGAGCGCTTCCGCTCCCGAGCGCCCTTTCAGCGCTCTGT 85
 QY 701 TCACAGCTGGGAGACGACGCGCTTCC 727
 DB 84 TCTTGGCAGTGAATCTGCGCGCATCC 58

RESULT 11
 CNS0175Y
 LOCUS
 DEFINITION CNS0175Y 1101 bp DNA linear GSS 26-JUL-1999
 Drosophila melanogaster genome survey sequence Sp6 end of BAC
 BACN37L08 of DrosBAC library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.

ACCESSION AL108460.1 GI:5628764
 VERSION
 KEYWORDS GSS.
 SOURCE Drosophila melanogaster.
 ORGANISM Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1101)
 Genoscope.
 Direct Submision
 Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 Evry cedex - FRANCE (E-mail : segr@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)

COMMENT
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the European Drosophila Genome Project (EDGP) -
 http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
 library (Dros BAC) was made by Alain Billaud at CEPH (Centre
 d'Etude du Polymorphisme Humain) with funding provided by a MRC
 project grant. The DNA was prepared from embryos from Alain Bucheton
 and Genevieve Payan. It has been constructed in the vector
 pBelBAC11.

FEATURES
 source
 1..1101
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_1lb="DrosBAC"
 /plasmid="pBelBAC11"
 /note="end : Sp6"

BASE COUNT 254 a 176 c 160 g 152 t 359 others
 ORIGIN

Query Match 2.9%; Score 43; DB 17; Length 1101;
 Best Local Similarity 14.4%; Pred. No. 1.3; Mismatches 132; Indels 2; Gaps 1;
 Matches 51; Conservative 169; Mismatches 169; Indels 2; Gaps 1;

QY 746 CACCTGGAATTATCTCTCACTGACGCGCATGAGGAGACAGCGACCCGCTGTGGCTC 805
 DB 711 MACSSSSASSSSGSCASTSSASRGMVSSCACSSGSGSAGSAGAGCGGGRGGG 770
 QY 806 TAGATGGAGTTTCATGCAATCCGCGCACTTCCAGCAGAAAGTCCCGCAATTACCGCT 865
 DB 771 GCCASSGVCAGACSSASCSSASMGVSSGSSGSCASCSCG--VSSCAVASASSSVMS 828
 QY 866 ACAGCGCTGCCAAGCCATACCCGCGAAGCGCGGGAACCTACCCAGAGAGAGTGT 925
 DB 829 KVASAVASCSAVASGMAAGAVSSSCRSSVASVMAAASVSSSVSSSVSAVAAS 888
 QY 926 CAGAGTCTCTGCGCGCGCATGTTCGCGCTTGAAGAGCGCGCGCGCATTCACCTCG 985
 DB 889 SSSASASMAVAAAAAVASVSVASVSSSSSCSSSSASVAVASVAVASVSSSS 948
 QY 986 CGCGAGCGCGGACGACCCAGCGCTCGCGCGCGACCGCAGAGAACACATTTGATT 1045
 DB 949 SSVETSSASVSVASVMSVAVSSASSSSVSSSVSVVAAAAAASSSSS 1008
 QY 1046 ACTAGACACCGCTACGAGCAGAGCGCGCTCCCTTCGCTGACACAGTGAAGA 1099
 DB 1009 ASAAVAVASSSSSSSSSSSSSSVSSSVSSSVSVSSSVSVVAVASA 1062

RESULT 12
 BE718105
 LOCUS
 DEFINITION BE718105 640 bp mRNA linear EST 12-SEP-2000
 RCI-HT0797-210600-021-b01 HT0797 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BE718105
 VERSION BE718105.1 GI:10106370
 KEYWORDS EST.
 SOURCE human.
 ORGANISM human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 640)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
 Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil

TEl: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l=RC1-HT0797-210
 600-021-b01&f=2000-06-21&t=1)
 Seq primer: puc 18 forward
 High quality sequence stop: 8
 High quality sequence stop: 579.

FEATURES
 source
 1..640
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

```

/dev_stages="Adult"
/clone_lib="HT0797"
/note="Organ: head, neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 127 a 210 c 179 g 124 t
ORIGIN

```

```

Query Match 2.9%; Score 42.6; DB 12; Length 640;
Best Local Similarity 47.3%; Pred. No. 1.3; Indels 0; Gaps 0;
Matches 129; Conservative 0; Mismatches 144;
QY 461 CCCAAACCTCGGGAGCAGCTCAGCAATGGTGGCGCCCTTTTACTGTGATCCACCATCC 520
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 352 CCATTTACCTGATGAAGATCTGACGGAGCGCGCTACAGCTTACACACCGCCGAGC 411
QY 521 ATGGTGTCTTTCTGATGTTAGCGACGAGCAGCTTTCTGGATCAGTTCCCGATGATA 580
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 412 GTGAGATCGTGGCGGACATCAAGGAGAAGCTGTGCTACGTGGCTCTGGACTTCGAGCAGG 471
QY 581 TCTCTTCGACCTTTGGTTCCTCTCATCTACTCAGGTGCGAGCGCTCAGTGGTCTCGCGGTC 640
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 472 AGATGGCCACCGCTGCGCGCTCCACCTCGCTGGAGAAGTCGTACGAGTTGCTTGACGGCC 531
QY 641 CAGGAGATCAGGCTGGTCTATGAACACTTTCGGTTCCTCGACGGCATCTCCGACCTCTCAG 700
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 532 AGTGATCACCATTGGCAACGAGCGCTTCGCTGCCCGAGGCCCTGTTTCAGACCTCGT 591
QY 701 TCACAGGCTGGGAGACGACCGTCTTCCCTGGAC 733
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 592 TCCTGGCGATGGAGTCGTGGCGCATCCGAGAC 624

```

```

RESULT 13
LOCUS AW600680 760 bp mRNA linear EST 12-JUN-2002
DEFINITION ESTPMC028 Penaeus monodon's total hemocyte cDNA library (#2)
Penaeus monodon cDNA clone PMC028 5' similar to actin, mRNA
sequence.
ACCESSION AW600680
VERSION AW600680.1 GI:11037809
KEYWORDS black tiger shrimp.
SOURCE Penaeus monodon
ORGANISM Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea;
Penaeidae; Penaeus.
REFERENCE 1 (bases 1 to 760)
AUTHORS Wongsantichon,J., Boonchooy,C., Udomkit,A., Panyim,S. and
Sonthayanon,B.
Expressed Sequence Tags from Black Tiger Prawn's Hemocytes
Unpublished (2000)
Contact: Sonthayanon, B.
Prawn Molecular Biology
Institute of Molecular Biology and Genetics, Mahidol University,
Salaya campus
Phutthamonthon 4 Rd., Phutthamonthon District, Nakhon Pathom, 73170
, Thailand
Tel: 66 2 4419003
Fax: 66 2 4419906
Email: scbat@mahidol.ac.th
Seq primer: T3.
FEATURES
Location/Qualifiers
1..760
/organism="Penaeus monodon"
/db_xref="taxon:6687"
/clone="PMC028"
/clone_lib="Penaeus monodon's total hemocyte cDNA library
(#2)"

```

```

/tissue_type="hemocytes"
/dev_stages="16-17 g average body weight (farm-raised)"
/lab_host="E.coli XL-1-Blue MRF"
/note="Vector: lambdaZAP II; Site_1: EcoR I; Site_2: Xho
I; The library was prepared using protocol given by
supplier, Stratagene, Inc."
BASE COUNT 163 a 238 c 162 g 176 t 21 others
ORIGIN
Query Match 2.8%; Score 42; DB 10; Length 760;
Best Local Similarity 47.4%; Pred. No. 2;
Matches 126; Conservative 0; Mismatches 140; Indels 0; Gaps 0;
QY 467 ACCTGGGACGACCTCAGCCATGGTGGCGCCCTTTTACTGGTACCACCATCCATGGTG 526
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 67 ACTGATGAAGATCTTCGACCGAGCGGTGTACACCTTCAGACACCGCGGAGGAGAA 126
QY 527 TCTTCTGATTTGGTAGCGACGAGCAGCTTCTTGGATCAGTTACGGATGATATCTCTT 586
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127 TCGTGGCCACATCAGGAGAAGCTGTGTAGTGGCTCTGGACTTCGAGCAGGAGATGA 186
QY 587 CGACCTTTTGGTTCCTCCATCCTCAGGTGCGAGCGCTCAGTGGGTCTGCGCGTCCAGGAG 646
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 187 CCACCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 246
QY 647 ATCAGGCTGTATGAAACACTTCGGGTTCTCGACGGCATCTCGAGCGCTCTAGTCACAG 706
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 247 TCACCATCGGCAACGAAGGTTCCGCTGCCCGAGGCCCTGTTCACAGCCCTCATTCCTCG 306
QY 707 GCTGGGAGACGACCGCTTCCTCCCTGGA 732
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 307 GCATGGAGTCGTGGCGCATCCAGAA 332

```

```

RESULT 14
LOCUS CNS016BR 1201 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence 17 end of BAC
BACN15K14 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL106545
VERSION AL106545.1 GI:5622456
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk . This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CESP (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.
FEATURES
Location/Qualifiers
1..1201
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACN15K14"
/clone_lib="DrosBAC"
/plasmid="pBelobAC11"
/note="end : T7"
BASE COUNT 386 a 182 c 234 g 101 t 298 others
ORIGIN

```

Query Match	2.8%	Score 41.8;	DB 17;	Length 1201;
Best Local Similarity	21.4%	Pred. No. 3;		
Matches	44;	Conservative	85;	Mismatches 77; Indels 0; Gaps 0;

QY	824	CATTCCGGCACTTCCAGCAGAGAGTTCGCCGAATTCAACGCGTACACGCTCGCCAA	CGCA	883
Db	820	CMYTTCCTCCMACMCMCMAMMCCCTTCTKAGTANCTCMAMMMCSGGSCAAMVVV		879
QY	884	TACCCGGCAACAGCGCGGGAACCTCACCCACAGAAAGTGCAGAGTCTCTCGGCGGC		943
Db	880	VVVVVVVVVSYCCCCMAGMACCHTATATTCGCMGBGASCGBSVSSGGSCVRRSSVGR		939
QY	944	GCATGTTCCGCGCGTTGAGAGACGCGCGCGCGCATTTGACTCGCGCGGACGCGGAC		1003
Db	940	MYHTYGGKBCRGRGTTWVSCSSGGSSSCGKTVGSCSSSCSGGSCCRBCTCGSS		999
QY	1004	CAGCGCTGCGCGCCGACCCCGACGAG		1029
Db	1000	SSSSSCBSGTSCTTYSSSSSSSSK		1025

LOCUS	DEFINITION	EST
BU309091/c	615 bp mRNA linear	EST-09-APR-2002
BU309091	Y. Ogihara unpublished cDNA library, Wh_Yd	
BU309091	aestivum cDNA clone whYd1813 3', mRNA sequence.	

ACCESSION	BJ309091
VERSION	BJ309091.1
	GI:20117255

KEYWORDS	EST.
SOURCE	bread wheat

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae

REFERENCE 1 ; Trillicae; Trillicum.
(bases 1 to 615)

AUTHORS	Ogihara, Y. and Murai, K.
TITLE	Expressed genes in <i>Triticum aestivum</i>

JOURNAL	Unpublished (2002)
COMMENT	Contact: Tadasu Shin-i

Center For Genetic Resource Information
National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-5559-81-6856

Fax: 81-559-81-6855
Email: tehini@enes.nic.ac.ir

FEATURES	Location/Qualifiers
COURSE	1 61E

```
1.015
/organism="Triticum aestivum"
source
```

```

/cultivar="Chinese Spring"
/db_xref="taxon:4565"

```

```

/clone="whyd18113"
/clone_lib="Y. Ogiwara unpublished cDNA library, wh_yd"

```

```
/tissue_type="spikelet at late flowering"  
/dev_stage="Feekes' scale 6"
```

```
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;  
Site 1: EcoRI; Site 2: XhoI; Plants were grown under
```

hydroponic conditions at UC Davis, salt stressed for 12 hours, and for 7 days, then dissected and frozen (Akhunov

in J Dvorak Lab). Total RNA was prepared from sheath tissue, equal quantities of RNA were pooled from the two

samples, polyA was purified from the pooled RNA, a cDNA library was made, and the cDNA clones were in vivo

excised to give pBluescript phagemids in the TJ Clouse lab at the University of California, Riverside (Akhmoy, Chin

, Choi, Close, Fenton, Kianian, Otto, Simons, Zhang). Plasmid DNA preparations and DNA sequencing were

BASE COUNT 146 a 168 c 139 d 162 f 1 others

ORIGIN	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	

Query Match 3 88: Score 41 4: DB 13: Tenth 615:

	Local Similarity	DB-LSI	TrendS	Cross
Query Match	2.08;	DB LSI	denglu 0.13;	
Best Local Similarity	54.24;	Pred. No. 2.6;		
Matches	94;	Mismatches	71;	0;

Qy	897	CGCGGAAAACCTCAACCGACGAGGAAGGTGCAAGTTCTCTGGGCGCGGCAATGTTGGGCGC	956
Db	506	CGAGAGAACTGGAGCCCGGAGCAAGTCTTCTTCTCGCGCGCTCTCTCAG	447
Qy	957	TTGAAAGCGCGCGCGCGATTTGACTCTGGGCGGACGCGGCGAAGCCACCGCTGGCGC	101
Db	446	CAGGAACAGGATCTGTGTTCTTCGACGAGGCGAAGCGCTCATGACTCGGCGCACTGACGC	387
Qy	1017	CGACCCGACGAGAAACAATTTGCATTACTAG	1051
Db	386	CATCTTCAGAGGTCATCAAGCAGGAGTTCTCAG	352

Search completed: December 4, 2002, 23:12:16
Job time : 1952 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2002, 14:49:33 ; Search time 59 Seconds
(without alignments)

7765.679 Million cell updates/sec

Title: US-09-926-084-8

Perfect score: 1494
Sequence: 1 atgcgccttcgcctgttgc.....tcactgcgactatcgcgtc 1494

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database:

Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTCUTS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43.4	2.9	20235	1 US-07-642-734C-3	Sequence 3, Appl1
2	43.4	2.9	20235	3 US-08-439-009A-3	Sequence 3, Appl1
3	42.4	2.8	9757	1 US-08-093-453B-1	Sequence 1, Appl1
4	42.4	2.8	9759	1 US-08-459-041A-1	Sequence 1, Appl1
5	42.4	2.8	9759	3 US-08-999-713-1	Sequence 1, Appl1
6	40.2	2.7	13613	4 US-09-105-537-3	Sequence 3, Appl1
7	38.8	2.6	2167	2 US-08-461-775-9	Sequence 9, Appl1
8	38.8	2.6	2167	2 US-09-031-606-9	Sequence 9, Appl1
9	38.8	2.6	2668	2 US-08-461-775-11	Sequence 11, Appl1
10	38.8	2.6	2668	2 US-09-031-606-11	Sequence 11, Appl1
11	37.2	2.5	538	4 US-08-998-416-1056	Sequence 1056, Ap
12	37	2.5	796	4 US-08-998-416-467	Sequence 467, App
13	36.8	2.5	1248	4 US-09-105-537-7	Sequence 7, Appl1
14	36.8	2.5	5970	3 US-09-320-878-21	Sequence 21, Appl1
15	36.6	2.4	68750	3 US-09-335-409-1	Sequence 1, Appl1
16	36.6	2.4	68750	4 US-09-568-102-1	Sequence 1, Appl1
17	36.6	2.4	68750	4 US-09-567-969-1	Sequence 1, Appl1
18	36.6	2.4	68750	4 US-09-568-480-1	Sequence 1, Appl1
19	36.6	2.4	68750	4 US-09-568-486-1	Sequence 1, Appl1
20	36.6	2.4	68750	4 US-09-568-472-1	Sequence 1, Appl1
21	36.6	2.4	68750	4 US-09-567-899-1	Sequence 1, Appl1
22	35.6	2.4	11220	4 US-09-105-537-32	Sequence 32, Appl1
23	35.6	2.4	35081	2 US-08-752-760A-1	Sequence 1, Appl1
24	35.6	2.4	36778	4 US-09-105-537-5	Sequence 5, Appl1
25	35.6	2.4	38506	3 US-09-320-878-19	Sequence 19, Appl1
26	35	2.3	1531	1 US-08-449-986-1	Sequence 1, Appl1
27	35	2.3	1531	2 US-08-756-855-1	Sequence 1, Appl1

28	34.8	2.3	43280	2 US-08-804-227C-1	Sequence 1, Appl1
29	34.6	2.3	2329	4 US-09-411-977-1	Sequence 1, Appl1
30	34.4	2.3	1519	1 US-08-225-477B-2	Sequence 2, Appl1
31	34.4	2.3	1519	5 PCT-US95-04353-2	Sequence 2, Appl1
32	34.4	2.3	11219	3 US-07-642-734C-1	Sequence 1, Appl1
33	34.4	2.3	11219	3 US-08-439-009A-1	Sequence 1, Appl1
34	34.2	2.3	758	2 US-08-288-630-2	Sequence 2, Appl1
35	34	2.3	2900	1 US-08-034-650-9	Sequence 9, Appl1
36	34	2.3	2900	1 US-08-449-015-9	Sequence 9, Appl1
37	33.8	2.3	666	2 US-08-875-034A-1	Sequence 1, Appl1
38	33.8	2.3	2299	4 US-09-153-599A-1	Sequence 1, Appl1
39	33.8	2.3	9757	1 US-08-093-453B-1	Sequence 1, Appl1
40	33.8	2.3	9759	1 US-08-459-041A-1	Sequence 1, Appl1
41	33.8	2.3	9759	3 US-08-999-733-1	Sequence 1, Appl1
42	33.6	2.2	30001	1 US-08-125-468-1	Sequence 1, Appl1
43	33.6	2.2	30001	2 US-08-474-933-1	Sequence 1, Appl1
44	33.6	2.2	53526	3 US-08-658-136-2	Sequence 2, Appl1
45	33.6	2.2	53577	3 US-08-658-136-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-07-642-734C-3
Sequence 3, Application US/07642734C
Patent No. 5824513
GENERAL INFORMATION:
APPLICANT: Katz, L
APPLICANT: Donadio, S
APPLICANT: McAlpine, J B
TITLE OF INVENTION: Recombinant DNA Method for Producing
TITLE OF INVENTION: Erythromycin Analogs
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Edward H. Gorman
STREET: Abbott Laboratories D377/APeD-2 One Abbott
CITY: Abbott Park
STATE: IL
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/642,734C
FILING DATE: 17-JAN-91
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Danckerts, Andreas M
REGISTRATION NUMBER: 32652
REFERENCE/DOCKET NUMBER: 4952.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-9396
TELEFAX: 708-938-2623
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 20235 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 238
FEATURE:
NAME/KEY: CDS
LOCATION: 19..10722

```

; OTHER INFORMATION: /codon_start= 19
; OTHER INFORMATION: /function= "gene_eryA"
; OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for
; OTHER INFORMATION: 6-deoxyerythronolide B"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19..4470
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 3"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 97..1482
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1693..2670
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyltransferase domain module 3"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3406..3921
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoreductase domain of module 3"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 4171..4428
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 3"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 4471..10722
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 4"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 6054..7026
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyltransferase domain of module 4"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7165..9216
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: dehydratase and enoylreductase domains m"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 9433..9984
; OTHER INFORMATION: /function= "approximate span
; OTHER INFORMATION: beta-ketoreductase of module 4"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10225..10483
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 4"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 10723..20235
; OTHER INFORMATION: /codon_start= 10723
; OTHER INFORMATION: /function= "gene_eryA"
; OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6
; OTHER INFORMATION: 6-deoxyerythronolide B formatio"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10723..15165
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 5"
; FEATURE:
; NAME/KEY: misc_feature

```

```

; LOCATION: 10831..12174
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 12379..13350
; OTHER INFORMATION: /function= "approximatr span of
; OTHER INFORMATION: acyltransferase domain of module 5"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 14062..14610
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoreductase of module 5"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 14857..15114
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 5"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15166..20235
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 6"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 15172..16569
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 16768..17721
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyltransferase domain of module 6"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 18379..18921
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoreductase domain of module 6"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19149..19398
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 6"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19492..20235
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: thioesterase domain of module 6"
;
; US-07-642-734C-3
;
; Query Match 2.9%; Score 43.4; DB 1; Length 20235;
; Best Local Similarity 46.1%; Pred. No. 0.018;
; Matches 182; Conservative 0; Mismatches 211; Indels 2; Gaps 1;
;
; QY 845 AGGTCCCGCAATTCAACGCGTACACGCTCGCCACACCGGATACCCGCGAACAGCGCGGAA 904
; DB 12623 AGCCGCCCGCGTCTGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGACGTGCGCCGCG 12682
; QY 905 ACCTCACCCAGCAGGAAGGTGCGAGAGTTCTCGCGCGCGGCATGTTCCGGCCGTTGGAAGA 964
; DB 12683 CGCTCACCTTGGAGACGCGCCNAGCTGTCGGTTCGGAGCCGGTCTCTGGCCCGGC 12742
; QY 965 GCGGCGCGCGATTGACTTCGCGCCGACCGCGGACGACCCAGCGTTCGCGCGCGACCCGC 1024
; DB 12743 TCGCGCGCGAGCGCGCATGGCGTCTTTCGGCTGGGCACCGAGCAGCGCGCGCGAACGGA 12802
; QY 1025 AGAGGACAACAATTTTCGATTACTCAGACACGCTGACGACGACGACGCGCTGCCCTTCG 1084
; DB 12803 TCGGCGCGTTCGCGCGCGCGCTCTCCATCGCTCGGTCAACGCGCCCGGTCGGTCTCG 12862
; QY 1085 GTGCACACGTCGAGAGACGAAACCCCTCGACAGGACCTGGGTGGACCGGTTCGACACCTT-- 1142
; DB 12863 TCGCGGGGAGAGCGGCGCGCTGGAGCGAGCTGATCGCCAGTGCAGGCGCGAAGGATAA 12922

```

QY 1143 CCAAGCATATCGGTTCCAGATCCCGACGGCCCAAAACGTTATGCACAACTTGCGTC 1202

Db 12923 CCGGGGGCGCGCATCCCGTCGACTAGCGCTCCCACTCAACCGCAGTGAATCGCTGGCG 12983

QY 1203 GGGCGTGACTGCGCAAGACCGCGGCTTTCTTTTTCG 1237

Db 12983 AGGAGCTGCTGACCGCAGCTGGCGGGCATTTTCCCG 13017

RESULT 2

```

US-08-439--009A-3
; Sequence 3, Application US/08439009A
; Patent No. 6004787
;
GENERAL INFORMATION:
APPLICANT: Donadio, S
APPLICANT: Katz, L
APPLICANT: Mcalpine, J B
TITLE OF INVENTION: Method of Directing Biosynthesis of
TITLE OF INVENTION: Specific Polyketides
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Steven F. Weinstein
STREET: Abbott Laboratories D377/AP6d-2 One Abbott
STREET: Park Rd
CITY: Abbott Park
STATE: IL
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/439.009A
FILING DATE: 11-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Caeulo, Dianne
REGISTRATION NUMBER: 40, 943
REFERENCE/DOCKET NUMBER: 4952.US.D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 847-938-3137
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2025 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 238
FEATURE:
NAME/KEY: CDS
LOCATION: 19..10722
OTHER INFORMATION: /codon_start=19
OTHER INFORMATION: /function="gene eryA"
OTHER INFORMATION: /product="eryA ORF2 encoding modules 3 & 4 for
OTHER INFORMATION: 6-deoxyerythronolide B"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 19..4470
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: module 3"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 97..1482
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketacyl ACP synthase of module 3"

```

```

1 FEATURE:
2 NAME/KEY: misc_feature
3 LOCATION: 1693..2670
4 OTHER INFORMATION: /function= "approximate span of
5 OTHER INFORMATION: acyltransferase domain module 3"
6 FEATURE:
7 NAME/KEY: misc_feature
8 LOCATION: 3406..3921
9 OTHER INFORMATION: /function= "approximate span of
10 OTHER INFORMATION: beta-ketoreductase domain of module 3"
11 FEATURE:
12 NAME/KEY: misc_feature
13 LOCATION: 4171..4428
14 OTHER INFORMATION: /function= "approximate span of
15 OTHER INFORMATION: acyl carrier domain of module 3"
16 FEATURE:
17 NAME/KEY: misc_feature
18 LOCATION: 4471..10722
19 OTHER INFORMATION: /function= "approximate span of
20 OTHER INFORMATION: module 4"
21 FEATURE:
22 NAME/KEY: misc_feature
23 LOCATION: 4471..5847
24 OTHER INFORMATION: /function= "approximate span of
25 OTHER INFORMATION: beta-ketocacylACP synthase domain of module 3"
26 FEATURE:
27 NAME/KEY: misc_feature
28 LOCATION: 6054..7026
29 OTHER INFORMATION: /function= "approximate span of
30 OTHER INFORMATION: acyltransferase domain of module 4"
31 FEATURE:
32 NAME/KEY: misc_feature
33 LOCATION: 7165..9216
34 OTHER INFORMATION: /function= "approximate span of
35 OTHER INFORMATION: dehydrogenase and enoylreductase domains m"
36 FEATURE:
37 NAME/KEY: misc_feature
38 LOCATION: 9433..9984
39 OTHER INFORMATION: /function= "approximate span
40 OTHER INFORMATION: beta-ketoreductase of module 4"
41 FEATURE:
42 NAME/KEY: misc_feature
43 LOCATION: 10225..10483
44 OTHER INFORMATION: /function= "approximate span of
45 OTHER INFORMATION: acyl carrier domain of module 4"
46 FEATURE:
47 NAME/KEY: CDS
48 LOCATION: 10723..20235
49 OTHER INFORMATION: /codon_start= 10723
50 OTHER INFORMATION: /function= "gene =eryA"
51 OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6
52 OTHER INFORMATION: 6-deoxyerythronolide B formatio"
53 FEATURE:
54 NAME/KEY: misc_feature
55 LOCATION: 10723..15165
56 OTHER INFORMATION: /function= "approximate span of
57 OTHER INFORMATION: module 5"
58 FEATURE:
59 NAME/KEY: misc_feature
60 LOCATION: 10831..12174
61 OTHER INFORMATION: /function= "approximate span of
62 OTHER INFORMATION: beta-ketocacylACP synthase domain of modu1"
63 FEATURE:
64 NAME/KEY: misc_feature
65 LOCATION: 12379..13350
66 OTHER INFORMATION: /function= "approximatr span of
67 OTHER INFORMATION: acyltransferase domain of module 5"
68 FEATURE:
69 NAME/KEY: misc_feature
70 LOCATION: 14062..14610
71 OTHER INFORMATION: /function= "approximate span of
72 OTHER INFORMATION: beta-ketoreductase of module 5"
73 FEATURE:

```

```
; NAME/KEY: misc feature
; LOCATION: 14857..15114
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 5"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 15165..20235
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 6"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 15172..16569
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketocylACPsynthase domain of modul"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 16768..17721
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyltransferase domain of module 6"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 18379..18921
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoreductase domain of module 6"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 19149..19398
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 6"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 19492..20235
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: thioesterase domain of module 6"
; US-08-439-009A-3
```

```
Query Match 2.9%; Score 43.4; DB 3; Length 20235;
Best Local Similarity 46.1%; Pred. No. 0.018;
Matches 182; Conservative 0; Mismatches 211; Indels 2; Gaps 1;

QY 845 AGGTCCTCCCGAATTCAACGCGTACACGCTCGCCCAACGCGATACCCGCAACAGCGCGGAA 904
DB 12623 AGCCCGCGCGGTCTCGGCCACTCGCAGGCGAGATCGCGCGCGCACGTGCGCGCG 12682

QY 905 ACCTCACCCAGCAGGAGTTCAGAGTTCTCTGCGCGCGCGATGTTCTGGCGGTTGGAAGA 964
DB 12683 CGCTCACCTTGGAGACGCGCGCATGCTGCGGCTCGGAGCGCGGTCTGGCCCGGC 12742

QY 965 CGCGCGCGCGATTGACTCGCGCGCGCGCGGAGCAGCAGCGCTCGCGCGCGACCCGC 1024
DB 12743 TCGCGCGCGCGCGCGGATGCGTCTGCGGCTGGGCGCAGCAGCGCGCGCGAACGGA 12802

QY 1025 AGAGGAACAACATTTTCGATTACTCAGACACGCTGACGAGCAGCAGCGCTGCCCCCTTCG 1084
DB 12803 TCGCGCGCTTTCGCGGCGCGCTCTTCATCTGCTCGCTCGTCAACGCGCGCGCTGCTG 12862

QY 1085 GTGCACACGTGAGGAAGAACACCTTCGACAGGACCTTGGTGGACCGGTGCACACCTT-- 1142
DB 12863 TCGCGGGGAGAGCGGCGCGCTGGACGAGCTGTATCGCGGAGTGGCGAGCGCGAAGGCATAA 12922

QY 1143 CCAGCTATCGGTTCAGTATCCGTTACGCGCCGAGAACGCTGTGATCAGAACTTGGTC 1202
DB 12923 CGCGCGCGCGCATCCCGCTCGACTACGCTCCCACTCACCGCAGGTGGAGTCGCTGCGCG 12982

QY 1203 GGGCGTGACTTCGCAACACCGCGTCTTCTTTTCG 1237
DB 12983 AGAGAGTGTGACCGACTGCGGGGCATCTCCCGC 13017
```

```
RESULT 3
US-08-093-453B-1
; Sequence 1, Application US/08093453B
; Patent No. 5439814
```

```
; GENERAL INFORMATION:
; APPLICANT: Frey, Teryl K.
; APPLICANT: Dominguez, Geraldina
; APPLICANT: Wang, Chin-Yen
; TITLE OF INVENTION: Modified Infectious Rubella Virus
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jamie L. Greene, Jones & Askew
; STREET: 191 Peachtree Street, 37th Floor
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: United States
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50
; COMPUTER: Macintosh
; OPERATING SYSTEM: 7.0
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/093,453B
; FILING DATE: 19 JUL 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 07/722,334
; FILING DATE: 28 JUN 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Greene, Jamie L.
; REGISTRATION NUMBER: 32,467
; REFERENCE/DOCKET NUMBER: 07362-0101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404.818-3700
; TELEFAX: 404.818-3799
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9757 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: RNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Rubella virus
; STRAIN: Therien
; US-08-093-453B-1

Query Match 2.8%; Score 42.4; DB 1; Length 9757;
Best Local Similarity 47.8%; Pred. No. 0.025;
Matches 86; Conservative 8; Mismatches 86; Indels 0; Gaps 0;

QY 841 CAGAAAGTCCCCGAATTCAACGCGTACACGCTCGCAACGCGATACCCGCAACAGCGCG 900
DB 5600 CAUACGAGCGCGGAGGUCGAUGCGUGGUGGAGGCCAUUACACCAACGCAUCGAG 5659

QY 901 GGAACCTACCCAGCAGGAGGTGTCAGAGTTCTCTCGGCGCGCGCATGTTCTGCGCGTTGG 960
DB 5660 GUCGACUUCACUGAGUUCGACAUAAACACGACCCUCGCUACUCGCGGACGUCGAG 5719

QY 961 AAGAGCGGCGCGCGATTGACCTCGCGCGCGAGCCGCGGAGCAGCCAGCGTTCGCGCGCGAC 1020
DB 5720 AUUAGCGCGCUCUCUUGGGCCUCCUUGGCGCGGAAGAUACCGCGCGGCGCGCGCG 5779

RESULT 4
US-08-459-041A-1
; Sequence 1, Application US/08459041A
; Patent No. 5683065
; GENERAL INFORMATION:
; APPLICANT: Frey, Teryl K.
; APPLICANT: Dominguez, Geraldina
; APPLICANT: Wang, Chin-Yen
; TITLE OF INVENTION: DNA Encoding Infectious Rubella Virus
; NUMBER OF SEQUENCES: 11
```


Db 1878 GGGCGCGTCCGCGGATCAAGCGCCGAGATGAGTCGACCGACTGCGAGCTGGGAC 1931

RESULT 9

US-08-461-775-11
; Sequence 11, Application US/08461775
; Patent No. 5858773
; GENERAL INFORMATION:
; APPLICANT: MAZODIER, Philippe
; APPLICANT: GUGLIEMI, Gerard
; TITLE OF INVENTION: REGULATORY NUCLEOTIDE SEQUENCE OF THE
; TITLE OF INVENTION: INITIATION OF TRANSCRIPTION
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,775
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/050,313
; FILING DATE: 10-MAY-1993
; APPLICATION NUMBER: FR 9011186
; FILING DATE: 10-SEP-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feury, Sharon E
; REGISTRATION/DOCKET NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 010830-035
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2668 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-461-775-11

Query Match 2.6%; Score 38.8; DB 2; Length 2668;

Best Local Similarity 47.9%; Pred. No. 0.16; Matches 112; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

Db 832 CACTTCAGCAGAGGTCCTCCGAATTCAACGCTGACGCTCGCCAAAGGATACCGCG 891
Db 1698 GCGCGCAAGCGGATGCTCGGCGACATGCGCCACCTCAGCGGTGCGCATGCGCGAG 1757
Qy 892 AACAGCGGGGAAACCTCACCAGCAGAGAGTGCAGATTCTCGGCGCGCATGTTG 951
Db 1758 GAGGTGCGCTCAAGCTGACAGAGCGCGGTCTGAGCGTGTGGGACCGCGCGGCTC 1817
Qy 952 GGGCGTTGAAGAGCGGCGCGCGCATTTGACTCGCGCGAGCGGAGAGACCCAGCGTTC 1011
Db 1818 ACCGTACCAAGAGCAGACACACCATCTGTGACCTGGAAGAGAGCGCGAGAGCTCCAG 1877
Qy 1012 GGGCGCGACCGCGCAGAGAGAACATTTGATTACTGACAGCGCTGACCGGAC 1065
Db 1878 GGGCGCGTCCGCGGATCAAGCGCCGAGATCGAGTCGACCGACTGCGAGCTGGGAC 1931

RESULT 10

US-09-031-606-11
; Sequence 11, Application US/09031606
; Patent No. 6153404
; GENERAL INFORMATION:
; APPLICANT: MAZODIER, Philippe
; APPLICANT: GUGLIEMI, Gerard
; TITLE OF INVENTION: REGULATORY NUCLEOTIDE SEQUENCE OF THE
; TITLE OF INVENTION: INITIATION OF TRANSCRIPTION
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/031,606
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/050,313
; FILING DATE: 10-MAY-1993
; APPLICATION NUMBER: FR 9011186
; FILING DATE: 10-SEP-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feury, Sharon E
; REGISTRATION/DOCKET NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 010830-035
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2668 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-031-606-11

Query Match 2.6%; Score 38.8; DB 3; Length 2668;

Best Local Similarity 47.9%; Pred. No. 0.16; Matches 112; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

Qy 832 CACTTCAGCAGAGGTCCTCCGAATTCAACGCTGACGCTCGCCAAAGGATACCGCG 891
Db 1698 GCGCGCAAGCGGATGCTCGGCGACATGCGCCACCTCAGCGGTGCGCATGCGCGAG 1757
Qy 892 AACAGCGGGGAAACCTCACCAGCAGAGAGTGCAGATTCTCGGCGCGCATGTTG 951
Db 1758 GAGGTGCGCTCAAGCTGACAGAGCGCGGTCTGAGCGTGTGGGACCGCGCGGCTC 1817
Qy 952 GGGCGTTGAAGAGCGGCGCGCGCATTTGACTCGCGCGAGCGGAGAGACCCAGCGTTC 1011
Db 1818 ACCGTACCAAGAGCAGACACACCATCTGTGACCTGGAAGAGAGCGCGAGAGCTCCAG 1877
Qy 1012 GGGCGCGACCGCGCAGAGAGAACATTTGATTACTGACAGCGCTGACCGGAC 1065
Db 1878 GGGCGCGTCCGCGGATCAAGCGCCGAGATCGAGTCGACCGACTGCGAGCTGGGAC 1931

RESULT 11

US-08-998-416-1056
; Sequence 1056, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philipseen, Peter

APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtle, Philipp
APPLICANT: Rebischung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPHII
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESS: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8587
INFORMATION FOR SEQ ID NO: 1056:
SEQUENCE CHARACTERISTICS:
LENGTH: 538 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE: PAG1643UP
ORGANISM: PAG1643UP
US-08-998-416-1056

Query Match 2.5%; Score 37.2; DB 4; Length 538;
Best Local Similarity 45.5%; Pred. No. 0.24;
Matches 132; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

QY 895 AGCGCGGAAACCTCACCCAGCAGGAAGGTGCAGAGTTCTTCGCGCGCGCATGTTTCGC 954
DB 113 AGCGCGCAGGCTGGCGAGCAGGCGCCGCAAAATCGTACTACCGCTGGTCCGGAC 172

QY 955 CGTTGGAAGAGCGCGCGCGATTGACCTCGCGCGAGCGGGAGCAGCAGCGCTCGC 1014
DB 173 GCGCACAGCTCACCGCGCCACTGCTCGCGGTGTCCACGCGCGCATGACGCGGCTC 232

QY 1015 GCGACCGCAGAGGACACAAATTTTCGATTACTACACACGCTGACGAGCAGACGCGC 1074
DB 233 TACCGCTACACAGCAGATCAGCAAGTCGTTTCAGGACACCTGATCTACTGCGCCGCGC 292

QY 1075 TGCCCTTCGTGTCACACGTGAGGAAGCAACCTTCAGCAGGACCTGGGTGGACCGGTC 1134
DB 293 CGCTGCTGAGCAAGATCGAGCTGACCGAGTGTCTACAGCTGGACATGCTGCTGCTATG 352

QY 1135 GACACCTTCACGCTATGCGGTTCAGTATCCGTTACCGGCCCAAGAGCTC 1184
DB 353 GAGCAGCAGCAGGAGGCGCCAGCGAGTGTCAAGTTCAACCCATATACGTC 402

RESULT 12
US-08-998-416-467

Sequence 467, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippson, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtle, Philipp
APPLICANT: Rebischung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPHII
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESS: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8587
INFORMATION FOR SEQ ID NO: 467:
SEQUENCE CHARACTERISTICS:
LENGTH: 796 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE: PAG1334UP
ORGANISM: PAG1334UP
US-08-998-416-467

Query Match 2.5%; Score 37; DB 4; Length 796;
Best Local Similarity 49.7%; Pred. No. 0.32;
Matches 94; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 861 CGCGTACAGCTCGCCCAACGCGATACCCCGCAACAGCGCGGGAACCTCACCCAGCAGA 920
DB 262 CGCAGCCCGCTCGGCGTGACCTTCGCGCGGCTCCGCCCTGTTCTGCGCGCGCTCGC 321

QY 921 AGGTGCAGAGTTCTTCGCGCGCGCATGTTTCGCGCGTTGGAAGAGCGCGCGGATTGA 980
DB 322 CGTGCAGGCGCTGATCGCGCGGCTTCTCTGTTGGCGCGGACCGCGCGCAACGCGCTCGC 391

QY 981 CTTTCGCGCGCAGCGCGAGCAGCCAGCGCTCGCGCGCGCGCGCGAGGAACAACTTT 1040
DB 382 CGCGCGCGCGCGCTCGCGTGCTGCGCGGTAGCGCGCGCGGTGCGCAGAACGACAAGT 441

QY 1041 CGATTACTC 1049
DB 442 CGTCTGCGC 450

RESULT 13
US-09-105-537-7

```
; Sequence 7, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 1248
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; US-09-105-537-7
```

```
Query Match 2.5%; Score 36.8; DB 4; Length 1248;
Best Local Similarity 46.2%; Pred. No. 0.45;
Matches 158; Conservative 0; Mismatches 182; Indels 2; Gaps 1;
```

```
Qy 727 CTTGACAGCGGCTGCTCCCACTGGAATTACTCTCACTGAGCGGATGGGAGACGGGC 786
Db 63 CTTGAGGGGGGCGCCAAACCGCATCGACCGCGCTGTAAGAGCGGCTCGACCGGGC 122
Qy 787 ACCGACCGCTGTGGGCTCTAGATGGAGATTTCATGCGATTCCGGCACTTCCACAGAG 846
Db 123 CTTGACAGCGGCTGCTGCTCAACGGGG--GCCGCTGCTCCGAGTTGAGAGAGCGC 180
Qy 847 GTCCCGCAATTCAACGGGCTACAGCGCTCGCCAAACCGGATACCGCGGACAGCGGGAAC 906
Db 181 GTGCGCGGGCTCGCGGGGTCCGGATGCGGTGACCTTCAACCGCCGCGGCTC 240
Qy 907 CTACCCGACGAGAGGTGACAGATTCTCTCGCGCGCGCATGTTGCGCGTTGGAAGAGC 966
Db 241 CAGCTCTCGGCGACGCGCGCGGCTTCAACGGGAGATCATGCGCTGATGACGTTT 300
Qy 967 GCGCGCGCGATTGACCTTCGCGCCGACGCGGACGACCCAGCGCTCGGCGCGGAC 1026
Db 301 GCCCGCACCCCGACGACCTGCGGTGATCGGCTTCAACCGCGGCTTTCGCGCATCGAC 360
Qy 1027 AGAACAACAATTTCGATTACTCAGACACGCTGACGAGCGAG 1068
Db 361 CCGACACCGGCAACTTCGACCCGACCGAGGTGGCGCGCGC 402
```

RESULT 14

```
US-09-320-878-21/C
; Sequence 21, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
```

```
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 21
; LENGTH: 5970
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; US-09-320-878-21
```

```
Query Match 2.5%; Score 36.8; DB 3; Length 5970;
Best Local Similarity 46.2%; Pred. No. 0.88;
Matches 158; Conservative 0; Mismatches 182; Indels 2; Gaps 1;
```

```
Qy 727 CTTGACAGCGGCTGCTCCCACTGGAATTACTCTCACTGAGCGGATGGGAGACGGGC 786
Db 933 CTTGAGGGGGGCGCCAAACCGCATCGACCGCGCTGTAAGAGCGGCTCGACCGGGC 874
Qy 787 ACCGACCGCTGTGGGCTCTAGATGGAGATTTCATGCGATTCCGGCACTTCCACAGAG 846
Db 873 CTTGACAGCGGCTGCTGCTCAACGGGG--GCCGCTGCTCCGAGTTGAGAGAGCGC 816
Qy 847 GTCCCGCAATTCAACGGGCTACAGCTTCGCCAAGCGATACCGCGAAGCGCGGGAAC 906
Db 815 GTGCGCGGGCTCGCGGGGTCCGGATGCGGTGACCTGCAACCGCACGCGGCGCTC 756
Qy 907 CTACCCGACGAGAGGTGACAGATTCTCTCGCGCGCGCATGTTCCGCGGTTGGAAGAGC 966
Db 755 CAGCTCTCGCGACGCGCGCGGCTTACCGCGGAGATGATTCGCTGATGACGTTT 636
Qy 967 GCGCGCGCGATTGACCTTCGCGCCGACGCGGACGACCCAGCGCTCGGCGCGGAC 1026
Db 695 GCGCGCACCCCGACGACCTGCGGTGATGCGGTCAACCCCGGCTTTCGCGCATCGAC 636
Qy 1027 AGAACAACAATTTCGATTACTCAGACACGCTGACGAGCGAG 1068
Db 635 CCGACACCGGCAACTTCGACCCGACCGAGGTGGCGCGCGC 594
```

RESULT 15

```
US-09-335-409-1
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Lisgon, James
; APPLICANT: Moinar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTRILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-335-409-1
```

```
Query Match 2.4%; Score 36.6; DB 3; Length 68750;
Best Local Similarity 51.5%; Pred. No. 2.9;
Matches 84; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
```

```
Qy 925 GCAAGATTCTCGCGCGCCATGTTGCGCGCTTGAAGACGGCGCGCGATTGACCTC 984
Db 64309 GCCCGACGCTCTCGCAAGCTTGAAGCTTGGCGCGCTCTGTACACTTCGCGGCTC 64368
Qy 985 GCGCGACGGCGGACGACCCAGCGCTCGCGCGGACCGGACGAGAAACAATTTCAT 1044
Db 64369 GCTCGCGCGCGACATCGGCGCGCGCGCGCGCGCGCGCGAGCTTCGAGCGGACTCGCGGAT 64428
```

Qy 1045 TACTCAGACACGCTGACGGACGAGACGCGCTGCCCTTCGGTG 1087
Db 64429 GACGAGCCGACGAGGCCGAGGGCGTCCGCCGTTCCGCG 64471

Search completed: December 4, 2002, 23:18:00
Job time : 247 secs

QY 835 TTCCAGAGAGGTTCCCGAATTCAACCGGTACACGCTCGCCAAACGGGATACCGGGAAC 894
DB 1711 TGGACACAGGTGGTGGCGGAGGGCTGGCGAACAACCTTCGCAACGGGATGAGCAAGGAC 1770
QY 895 AGCGCGGAACCTCACCCAGCAGGAGGTGCAGAGTTCTCGCGCGCGCATGTTCCGC 954
DB 1771 GCCGTGGGACAGCGCACGGCGCGCGGTTCGGCCGCGCGGCTGGGATCTGCCGATG 1830
QY 955 CGTTGGAAGAGCGCGCGCGATTGACCTCGCGCGGACGCGGACGACCGAGCGCTCGGC 1014
DB 1831 TCCGGGAAACCGGCACACCGAGGCGCACGGTTCGCGCGCTTCGTGGGCTTCACCAAC 1890
QY 1015 GCCGACCCGAGAGGAACAATTTTCGATTACTCAGACACGCTGACGAGCAGACGCGC 1074
DB 1891 CGCTACGCGCGCGGAACACTACATCTACGAGCACTCCAGCTCGCGCAGACATCTGTGTTCC 1950
QY 1075 TGCCC 1079
DB 1951 GGCCC 1955

RESULT 2
US-09-860-846-3/C
; Sequence 3, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, Y.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 13613
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-3

Query Match 2.7%; Score 40.2; DB 9; Length 13613;
Best Local Similarity 52.0%; Pred. No. 0.043;
Matches 90; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
QY 971 CGCGGATTGACCTCGCGCGGACGACGCGCTCGCGGCGGACGCGCTCGCGGCGGACGAGGA 1030
DB 11725 CGCGAAGTACCGCGCGGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11666
QY 1031 ACAACAATTTTCGATTACTCAGACACGCTGACGAGACGCGCTGCGGCTCGGTCGAC 1090
DB 11665 TCCAGCGCCAGATTGACGCTCCCGATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11606
QY 1091 ACCTGAGGAAGAGCAACCTTCGACAGGACCTCGGTTGACGCGGCTCGGTCGACCTTC 1143
DB 11605 CCGAACGCTCGCGCGACCCCGGAGCCGATCTGGGGGCGGCGGTCCACGCGCGTC 11553

RESULT 3
US-09-861-289-3/C
; Sequence 3, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1

; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 13613
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-3
Query Match 2.7%; Score 40.2; DB 10; Length 13613;
Best Local Similarity 52.0%; Pred. No. 0.043;
Matches 90; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
QY 971 CGCGGATTGACCTCGCGCGGACGACGCGCTCGCGGCGGACGCGCTCGCGGCGGACGAGGA 1030
DB 11725 CGCGAAGTACCGCGCGGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11666
QY 1031 ACAACAATTTTCGATTACTCAGACACGCTGACGAGACGCGCTGCGGCTCGGTCGAC 1090
DB 11665 TCCAGCGCCAGATTGACGCTCCCGATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11606
QY 1091 ACCTGAGGAAGAGCAACCTTCGACAGGACCTCGGTTGACGCGGCTCGGTCGACCTTC 1143
DB 11605 CCGAACGCTCGCGCGACCCCGGAGCCGATCTGGGGGCGGCGGTCCACGCGCGTC 11553

RESULT 4
US-09-712-363-10
; Sequence 10, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 1584
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-10

Query Match 2.6%; Score 38.2; DB 9; Length 1584;
Best Local Similarity 51.5%; Pred. No. 0.055;
Matches 88; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
QY 846 GGTCCCCGGAATTCAACCGGTACACGCTCGCCCAACGGGATACCGCGGAGACGCGCGGAAA 905

Db 709 GGTATACCCGAGCAACGGTTACCCGAGCAACCGGCTTACCCGACCAAGCGGGGTAC 768
Qy 906 CCTCACCAGAGCAAGTGTGAGATTCCTCGGCGCGCATGTTCCGCGCTTGAAGAG 965
Db 769 CAGAACCAAGGTGAGGCTACCCGACCAAGGAGGAGGGGGCTTTCGCGCCCTACGAG 828
Qy 966 CGGCGCGCGCATTTGACCTCGGCGCGAGCGGCGAGCAACCGGCTCGGCGC 1016
Db 829 CAAGCGCCCTCTGTTTCTCCCGGCGCGGCTGCGGCTACGCGGCTCCCGGC 879

RESULT 5

US-09-917-800A-115
; Sequence 115, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; LENGTH: 467
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 AAB48563
US-09-917-800A-115

Query Match 2.5%; Score 38; DB 10; Length 467;
Best Local Similarity 57.6%; Pred. No. 0.032;
Matches 68; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 966 CGGCGCGCGCATTTGACCTCGGCGCGAGCGGCGAGCAACCGGCTCGGCGCGCA 1025
Db 336 CGCGCGCAAGAACAGGCTGAGTGAACCCGAGAACCGGTTGACGCGGAAGCGGCT 395
Qy 1026 GAGGAACAACAATTTGACTGACAGACGCTGACGAGAGAGCGGCTGCCCTTC 1083
Db 396 GATCGGCGCGCAAGTTGCGGCGAGCCCGGTGTGATCGGACATGAAGCACTGCGCTTC 453

RESULT 6

US-09-917-800A-73/c
; Sequence 73, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:

; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; LENGTH: 515
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 AAB18604
; NAME/KEY: misc feature
; LOCATION: (1)..(515)
; OTHER INFORMATION: n = a or c or g or t
US-09-917-800A-73

Query Match 2.5%; Score 38; DB 10; Length 515;
Best Local Similarity 57.6%; Pred. No. 0.034;
Matches 68; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 966 CGGCGCGCGCATTTGACCTCGGCGCGAGCGGCGAGCAACCGGCTCGGCGCGCA 1025
Db 380 CGCGCGCAAGAACAGGCTGAGTGAACCCGAGAACCGGTTGACGCGGAAGCGGCT 321
Qy 1026 GAGGAACAACAATTTGACTGACAGACGCTGACGAGAGAGCGGCTGCCCTTC 1083
Db 320 GATCGGCGCGCAAGTTGCGGCGAGCCCGGTGTGATCGGACATGAAGCACTGCGCTTC 263

RESULT 7

US-09-917-800A-1386
; Sequence 1386, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02

; PRIOR APPLICATION NUMBER: US 60/290,029
 ; PRIOR FILING DATE: 2001-05-11
 ; PRIOR APPLICATION NUMBER: US 60/290,645
 ; PRIOR FILING DATE: 2001-05-15
 ; PRIOR APPLICATION NUMBER: US 60/292,336
 ; PRIOR FILING DATE: 2001-05-22
 ; PRIOR APPLICATION NUMBER: US 60/295,798
 ; PRIOR FILING DATE: 2001-06-06
 ; PRIOR APPLICATION NUMBER: US 60/297,457
 ; PRIOR FILING DATE: 2001-06-13
 ; PRIOR APPLICATION NUMBER: US 60/298,884
 ; PRIOR FILING DATE: 2001-06-19
 ; PRIOR APPLICATION NUMBER: US 60/303,459
 ; PRIOR FILING DATE: 2001-07-09
 ; NUMBER OF SEQ ID NOS: 1740
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 1386
 ; LENGTH: 2455
 ; TYPE: DNA
 ; ORGANISM: Rattus norvegicus
 ; FEATURE:
 ; OTHER INFORMATION: Genbank Accession No. US20020119462A1 L16764
 US-09-917-800A-1386

Query Match 2.5%; Score 38; DB 10; Length 2455;
 Best Local Similarity 57.8%; Pred. No. 0.082;
 Matches 68; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
 QY 966 CGCGCGCCGATTGACCTCGCGCGGACGCCGAGCCGCGCTCGCGCGCGCACCGCA 1025
 DB 336 CGCGCCGCAAGCAGCGTGGCGTGAACCCGCGACACCGTGTTCGACGGAAGCGCT 395
 QY 1026 GAGGAACAACAAATTCGATTACTAGACACGCTGACGAGCAGACGCGTGCCTTC 1083
 DB 396 GATCGGCGCAAGTTCGGCGACCGGTGGTGCAGTCGACATGAAGCACTGGCCCTTC 453

RESULT 8
 US-09-294-093B-5229
 ; Sequence 5229, Application US/09294093B
 ; Patent No. US20010051335A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lalgudi, Raghunath, V.
 ; APPLICANT: Ito, Laura, Y.
 ; APPLICANT: Sherman, Bradley, K.
 ; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
 ; FILE REFERENCE: PL-0009 US
 ; CURRENT APPLICATION NUMBER: US/09/294,093B
 ; CURRENT FILING DATE: 1999-04-16
 ; PRIOR APPLICATION NUMBER: 60/082,567
 ; PRIOR FILING DATE: April 21, 1998
 ; NUMBER OF SEQ ID NOS: 6207
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 5229
 ; LENGTH: 286
 ; TYPE: DNA
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: Incyte ID No. US20010051335A1 700356053H1
 ; NAME/KEY: unsure
 ; LOCATION: 36, 44, 81, 89, 92
 ; OTHER INFORMATION: a, t, c, g, or other
 US-09-294-093B-5229

Query Match 2.5%; Score 37.2; DB 10; Length 286;
 Best Local Similarity 47.5%; Pred. No. 0.044;
 Matches 105; Conservative 0; Mismatches 116; Indels 0; Gaps 0;
 QY 634 GC CGCGTCGAGGATCAGGCTGGTGCATGAACATTCGCGGTTCCTCGACGCGCATCTCGCAG 693
 DB 55 GCGAGCTCGAGCTCACTGGAGCANGGCGACTNCGNCACTCGATGGGGGAACGAGTGG 114

QY 694 CCCTCAGTCACAGGCTGGGAGCAGACCGCTTCCCTCGAGAGCGGTCGTCCACCTGGA 753
 DB 115 TAGGAGAACACGAGAGGAGGAGGAGGATGGAGAGCGGCGCTCTGTATCTTGGC 174
 QY 754 ATTATCCTCAGTCGACGCGATGGGACACCGGACACCGGACCGCTCGTGGGCTCTAGATGGG 813
 DB 175 ATGGGTGTCTCTGGAGGAGGAGGAGGTCGCCGCAACGGCCCGCTCTGTATGCCCGGAGGC 234
 QY 814 AGTTTCATGGCATTCGGGCACTTCCAGCAGAGAGGTCCCGCA 854
 DB 235 TCCTTCGAGGCTTACAAGTACGTCGAAGACATTTGTCTCAA 275

RESULT 9
 US-09-860-846-7
 ; Sequence 7, Application US/09860846
 ; Patent No. US2002016472A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sherman, D.H.
 ; APPLICANT: Liu, H.
 ; APPLICANT: Xue, Y.
 ; APPLICANT: Zhao, L.
 ; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
 ; FILE REFERENCE: 600.438US1
 ; CURRENT APPLICATION NUMBER: US/09/860,846
 ; CURRENT FILING DATE: 2001-05-18
 ; PRIOR APPLICATION NUMBER: 09/105,537
 ; PRIOR FILING DATE: 1998-06-26
 ; NUMBER OF SEQ ID NOS: 43
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 7
 ; LENGTH: 1248
 ; TYPE: DNA
 ; ORGANISM: Streptomyces venezuelae
 US-09-860-846-7

Query Match 2.5%; Score 36.8; DB 9; Length 1248;
 Best Local Similarity 46.2%; Pred. No. 0.14;
 Matches 158; Conservative 0; Mismatches 182; Indels 2; Gaps 1;
 QY 727 CTTGGACAGGCGGTCTCCACCTGGAATTATCTCTCAGTCGAGCGGATGGGACACGGGC 786
 DB 63 CTTCTGGGGGCGGCCCAACCGCATCGACCGCGCAGGCTGTACGAGCGGCTCGACCGGC 122
 QY 787 ACCGACCGCTGTGGGCTCTAGATGGAGTTTCATGGCACTTCGGGCACTTCAGCAGAAG 846
 DB 123 CTTGACAGCAGTGGCTGTCCAAACGGCG--GCCCGCTCGTCCGCGAGTTCGAGGAGCGC 180
 QY 847 GTCCCGGAATTCACCGGTACAGCTCGCCACGCGATACCCGCAACGCGGAGGCGGNAAC 906
 DB 181 GTCCCGGCGCTCGCGGGGTCCGGCATGCGTGCCCACTGCAACGCCACGCGCGGCTC 240
 QY 907 CTCACCCAGCAGGAAGTGTGAGAGTTCTCTCGGCGCGCATGTTCTGGCGCTTGAAGAGC 966
 DB 241 CAGCTCTCTCGGACACGCGCGCGGCTCAGCGGGAAGTATCATGCGCTCGATGAGCTTC 300
 QY 967 GCGCGCGCATTTGACCTCGCGCGAGCGGCGGAGCAGCCAGCGCTCGGCGCGGAGCAG 1026
 DB 301 GCGCGCACCCCGCAGCAGCTGCGCTGGATCGGCTCACCCTCGGCTCTTCGCGGACATCGAC 360
 QY 1027 AGGAACAACAAATTTTCGATTACTCAGACAGCTGACGAGCAG 1068
 DB 361 CGGACACCGGCAACCTCGACCCGACCGAGGTGGCGCGCG 402

RESULT 10
 US-09-861-289-7
 ; Sequence 7, Application US/09861289
 ; Patent No. US20020110897A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sherman, D.H.
 ; APPLICANT: Liu, H.
 ; APPLICANT: Xue, Y.

APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600.438US1
CURRENT APPLICATION NUMBER: US/09/861,289
CURRENT FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 09/105,537
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 7
LENGTH: 1248
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-861-289-7

Query Match 2.5%; Score 36.8; DB 10; Length 1248;
Best Local Similarity 46.2%; Pred. No. 0.14;
Matches 158; Conservative 0; Mismatches 182; Indels 2; Gaps 1;

QY 727 CTTGAGAGGCGGTCTGCTCCACCTGGAATTCTCTCACTGAGCGGATGGAGACGGGC 786
DB 63 CTTGAGAGGCGGTCTGCTCCACCTGGAATTCTCTCACTGAGCGGATGGAGACGGGC 122
QY 787 ACCGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 846
DB 123 CTTGAGAGGCGGTCTGCTCCACCTGGAATTCTCTCACTGAGCGGATGGAGACGGGC 180
QY 847 GTCCCGCAATTCACGCGCTACACCTGCGCAACCGGATCCCGGACAGCGCGGAAAC 906
DB 181 GTCCCGCGGCTGCTGCTCCACCTGGAATTCTCTCACTGAGCGGATGGAGACGGGC 240
QY 907 CTCACCGACGAGAGGCTGCTGCTCCACCTGGAATTCTCTCACTGAGCGGATGGAGAC 966
DB 241 CAGTCTCTGCGGACGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
QY 967 GCGCGCGGATTCCTGCTGCTCCACCTGGAATTCTCTCACTGAGCGGATGGAGAC 1026
DB 301 GCGCGCGGATTCCTGCTGCTCCACCTGGAATTCTCTCACTGAGCGGATGGAGAC 360
QY 1027 AGGAACAACATTTGATTACTCAGACACGCTGACGAGCGAG 1068
DB 361 CCGGACACCGGCAACTGACCTCGACCGGACGAGGTGGCGCGCG 402

RESULT 11
US-09-917-800A-1527
Sequence 1527, Application US/09917800A
Patent No. US20020119462A1
GENERAL INFORMATION:
APPLICANT: Mendrick, Donna
APPLICANT: Johnson, Mark
APPLICANT: Castle, Arthur
APPLICANT: Elashoff, Michael
TITLE OF INVENTION: Molecular Toxicology Modeling
FILE REFERENCE: 44921-5038-US
CURRENT APPLICATION NUMBER: US/09/917,800A
CURRENT FILING DATE: 2001-07-31
PRIOR APPLICATION NUMBER: US 60/222,040
PRIOR FILING DATE: 2000-07-31
PRIOR APPLICATION NUMBER: US 60/222,880
PRIOR FILING DATE: 2000-11-02
PRIOR APPLICATION NUMBER: US 60/290,029
PRIOR FILING DATE: 2001-05-11
PRIOR APPLICATION NUMBER: US 60/290,645
PRIOR FILING DATE: 2001-05-15
PRIOR APPLICATION NUMBER: US 60/292,336
PRIOR FILING DATE: 2001-05-22
PRIOR APPLICATION NUMBER: US 60/295,798
PRIOR FILING DATE: 2001-06-06
PRIOR APPLICATION NUMBER: US 60/297,457
PRIOR FILING DATE: 2001-06-13

PRIOR APPLICATION NUMBER: US 60/298,884
PRIOR FILING DATE: 2001-06-19
PRIOR APPLICATION NUMBER: US 60/303,459
PRIOR FILING DATE: 2001-07-09
NUMBER OF SEQ ID NOS: 1740
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 1527
LENGTH: 1366
TYPE: DNA
ORGANISM: Rattus norvegicus
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020119462A1 Z27118
US-09-917-800A-1527

Query Match 2.4%; Score 36.4; DB 10; Length 1366;
Best Local Similarity 56.8%; Pred. No. 0.19;
Matches 67; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 966 CCGCGCGCGGATTCGCTGCGCGGCGGAGACCGGCTGCGCGGCGGCA 1025
DB 64 CCGCGCGGAGACGAGTGGGCTGGAACCGGACCGGATTCGAGCGGAGCGGCT 123
QY 1026 GAGGAACAACATTTGATTACTCAGACACGCTGAGGAGAGCGGCTGCTGCTGCT 1083
DB 124 GATGCGCGGAGATTCGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 181

RESULT 12
US-09-815-242-7991
Sequence 7991, Application US/09815242
Patent No. US2002061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl B.
APPLICANT: Zyckind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 1410
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7991
LENGTH: 2208
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(2208)
US-09-815-242-7991

Query Match 2.4%; Score 35.6; DB 10; Length 2208;
Best Local Similarity 45.4%; Pred. No. 0.45;
Matches 128; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

```

QY 879 CGCGATACCCGGAACAGCGCGGGAACCTCACCAGCAGGAAGGTGCAGAGTTCTCGG 938
DB 1521 CGCCACCATCCCGGTGATCGCTCGCCCTTGATCGTGGCCCTGGCCAAATGAGCATGACGTCAC 1580
QY 939 CGCGCGCATGTTCCGGCGTTGGGAAGAGCGCGCGCGGCGGATTCAGCTCGCGCGGACGGCGGA 998
DB 1581 CCCTGCTGGCGGTGCCGCGGTGATCGGCTGGCCATCGGCTTCGGTGGCAGTCCGT 1640
QY 999 CGACCCAGCGCTCGGCGCGCGCCGAGAGGAACAAATTTTCGATTACTCAGACAGCT 1058
DB 1641 GGTGGCGGACCTCATCACCGCGCTGTTTCATCATCATCGAGGACTCGCTGTCGATCGACGA 1700
QY 1059 GACGGAGCAGCGCTGCGCCCTTCGCTGTCACAGTGTAGGAAGACGACCTTCGACAGGA 1118
DB 1701 CTAGCTCGAGTGGCGGCGCACCTTGGCAGCGTGTAGGGGCTGACTATCCGTTACCGTGG 1760
QY 1119 CTTGGGTGGACCGTGCACACCTTCCACGCTATGCGGTCCAG 1160
DB 1761 CTTGCGGACCTGAGCGGCTGTGCATACCATCCCGTTCCAG 1802

RESULT 13
US-10-044-090-591
; Sequence 591, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 591
; LENGTH: 4071
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 358933.9
; NAME/KEY: unsure
; LOCATION: 3867-3869, 3875
; OTHER INFORMATION: a, t, c, g, or other
US-10-044-090-591

Query Match 2.4%; Score 35.6; DB 12; Length 4071;
Best Local Similarity 45.4%; Pred. No. 0.64;
Matches 128; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 897 CGCGGAAACCTCACCAGCAGGAAGGTGCAGAGTTCTTCGGCGCGCGCATGTTTCGGCCG 956
DB 114 CGCCGACGCTTCGGCTGTCAGCAGCAGCCACCGGCCCGCTCCCGAGAGCCAGCGCCGCG 173
QY 957 TTGGAAGAGCGCGCGCGCATTTGACCTTCGCGCGCAGCGCGGAGACCCAGCGGCTCGCGC 1016
DB 174 AGCGCGCAGCGCGCGCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCGCGC 233
QY 1017 CGACCCGAGGAGAAACAATTTTCGATTACTCAGACACGCTGACGAGCAGCAGCGCGCTG 1076
DB 234 TTCTCTCGCGCGCTGCGGCCATGCGGCGGCGCTGCTGAGCTGCGCGCGCGCGCGCG 293
QY 1077 CCCTTTCCGTCACAGTGTAGGAGAGACAACTTCGACAGGACCTTGGGTGGACCGGTGCA 1136
DB 294 CGCTCCCGCTCGCCCGCAGCCGCACTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 353
QY 1137 CACTTCCAGCTATGCGGTTCAGTATCCCGTACGCGCCGAGA 1178
DB 354 TTCTTTCCAGCGCGGAGCGCGCGCGCGCTTAGGAGGAGA 395

RESULT 14
US-09-860-846-32
```

```

; Sequence 32, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-32

Query Match 2.4%; Score 35.6; DB 9; Length 11220;
Best Local Similarity 49.0%; Pred. No. 1.1;
Matches 95; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 828 CCGGCACTTCCAGCAGAAGGTCCCCGAATTCAACGCGTACACGCTCGCCAAACCGGATACC 887
DB 2832 CGGGCTGTCCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2891
QY 888 CGCGAACAGCGCGGGAACCTCACCAGCAGGAAGGTGCAGAGTTCTTCGGCGCGCGCAT 947
DB 2892 GGTGCTCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 2951
QY 948 GTTCGCGCGCTTGAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1007
DB 2952 GGGCGACCGCGCAGCGCTCGCGCGCAGCTGGCGGAGGCCCTGGCGCGCGCGCGCGCGCG 3011
QY 1008 GCTCGCGCGCGGACC 1021
DB 3012 CGTCGACGGCGTCC 3025

RESULT 15
US-09-861-289-32
; Sequence 32, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-32

Query Match 2.4%; Score 35.6; DB 10; Length 11220;
Best Local Similarity 49.0%; Pred. No. 1.1;
Matches 95; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 828 CCGGCACTTCCAGCAGAAGGTCCCCGAATTCAACGCGTACACGCTCGCCAAACCGGATACC 887
DB 2832 CGGGCTGTCCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2891
```

Qy 888 CGCGAAGAGCGCGGAAACCTCACCCAGAGAAAGTGCAGATTCTCGGCGCGGCAT 947
Db 2892 GGTGCTCGCGCGCTGTCCGCGCGCGCGCGACCCGTACAGCTGAGCGTGTCCGCT 2951
Qy 948 GTTCGCGCGTTGGAAGAGCGCGCGCGCATTTGACTCGCGCGAGCGGAGAGACCCAGC 1007
Db 2952 GGGCGACCGGCAAGCGGCTCGCGCGGACGCTGGGCGAGGCCCTGGGCGCGCGGTGAGC 3011
Qy 1008 GCTCGGCGCGGACC 1021
Db 3012 CGTCGACGGGCTCC 3025

Search completed: December 4, 2002, 23:14:05
Job time : 99 secs

THIS PAGE BLANK (USPTO)